Molecular and Genome Evolution in the Malesian Slipper Orchids (*Paphiopedilum* section *Barbata*)

Yap Jing Wei

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

Statement of originality

I, Jing Wei Yap, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date: 21/09/15

Details of collaboration and publications:

This project was done in collaboration with Dr. Rene Smulders, Wageningen UR, Netherlands and Dr. Barbara Gravendeel, Naturalis Biodiversity Center, Netherlands

Acknowledgements

My deepest and sincere thanks go out to my supervisors Prof. Andrew Leitch, Dr. Ilia Leitch, Dr. Mike Fay and Dr. Lee Yung-I for their guidance and support throughout my PhD study.

I would also like to acknowledge the financial support from the Commonwealth Scholarship and the Emily Holmes Scholarship, without which this PhD would not have been possible.

I am very grateful to my mentors, Dr. Sven Buerki, Dr. Laura Kelly and Dr. Oriane Hidalgo, for their help with crucial analyses.

I would also express my thanks to Prof. Mark Chase, Dr. Philip Cribb and André Schuiteman for their advice and useful contacts.

Vital data for this thesis was provided by Dr. Rene Smulders, Dr. Barbara Graveendel and Shairul Izan.

Many thanks go out to my fellow lab mates Steven Dodsworth, Hannes Becher, Peter Day, Maite Guinard, Bhanubongcheewin, Wang Wencai for their company, camaraderie and stimulating discussions. These past four years would not have been the same without them.

My thanks also go to Dr. Ng Yan Peng, Dr. Dion Devey, Dr. Jim Clarkson, Dr. Jaume Pellicer, Robyn Cowan, Chris Ryan, Bala Kompalli, Edith Kapinos, Lazlo Csiba from the Royal Botanic Gardens, Kew, and Dr. Ruth Rose from Queen Mary University of London for all their help and assistance.

My overseas sample collection was kindly facilitated by Saw Leng Guan, Ong Poh Teck & Avelinah Julius from the Kepong Herbarium, Forest Research Institute, Malaysia, Tony Lamb and Remi Ripin & Skybyn Sumail from Kinabalu Parks, Joan Pereira from Sandakan Herbarium, Linus Gosuking and Steven Chew from Kipandi Park, Sabah, Malaysia and Siti Roosita Ariati, Ita Puspita, Elizabeth Handini & Vitri Garvita from Bogor Botanic Gardens, Bogor, Indonesia.

Many valuable samples were also contributed by Paul Phillips from Ratcliffe Orchids Ltd., Bert Klein, Günter Gerlach & Jutta Babczinsky from Munich Botanic Gardens, Germany, Boris Schlumberger from Herrenhausen Gardens, Germany, and Derek Jackson and Alan Burdis from the British *Paphiopedilum* Society.

Last but not least, I thank my family and friends for supporting me throughout my studies and also in my personal life.

ABSTRACT

Paphiopedilum section *Barbata* (Cypripedioideae: Orchidaceae) is an evolutionarily young and charismatic group of terrestrial orchids native to the Himalayas, Indochina and Malesia. It contains several interesting species complexes, variable chromosome numbers (2*n*=28-42) and genome sizes (2C=55-70 pg) with hybrid speciation suspected on the basis of morphological data. In Chapter 1, I introduce *Paphiopedilum* and review existing literature on the group. In Chapter 2, I ask: what are the evolutionary relationships within section *Barbata*? I answer this by sequencing four plastid (*ycf1*, *matK*, *psa-ycf3ex3* and *trnF(GAA)-ndhJ*) and two low copy nuclear gene (*Xdh* and *CHS*) regions. Analysing the phylogenetic signals revealed patterns of gene tree incongruence and geographical groupings that suggest historical and on-going hybridisation. In Chapter 3, I ask: what biogeographical processes are driving diversification of section *Barbata*? To answer this I estimate the age of section *Barbata* from a dated BEAST phylogeny of cloned *Xdh* sequences and analyse it in relation to the geographical history of Southeast Asia. This revealed that section *Barbata* arose c. 6.3 Mya (95% HPD range=4.0-8.8 Mya) and that diversification is primarily driven by hybridisation, vicariance and dispersal facilitated by glacial-interglacial cycles of sea-level fluctuations in SE Asia, and possibly chromosomal changes. In Chapter 4, I ask: what genomic changes are occurring in section *Barbata*? I approach this by characterising repetitive DNA sequences in representative taxa and analyse them against new genome size estimates and published chromosome numbers. The results show that *Paphiopedilum* genomes are comprised of 61.1-71.5% repetitive DNA, and 28.9-39.5% single or low-copy DNA that is possibly derived from ancient repetitive elements. These findings suggest that a low-rate of repetitive DNA removal, rather than proliferation of any particular family of repetitive element, is driving genome evolution in the group. Finally in Chapter 5, I present my hypotheses on speciation processes in *Paphiopedilum* and outline avenues for future work.

Contents

CHAPTER 1
LITERATURE REVIEW AND INTRODUCTION TO EXPERIMENTAL THESIS

Changes in chromosome numbers (Table 1.5) are associated with the divergence of many plant and animal species. However, the evolutionary significance of the phenomenon is still not known and there is a long-standing question in evolutionary biology regarding the question as to whether certain chromosomal changes can actually drive speciation events or simply reflect slow genetic drift following speciation (Butlin 1993; Coyne and Orr 2004).

The Malesian *Paphiopedilum* section *Barbata*, which is the focus of the work presented here, represents an interesting orchid model for studying plant karyotype evolution as there is considerable diversity (39 taxa including varieties), chromosome number (2*n*=2*x*=28-42) and genome size diversity (Cox et al. 1998; Chochai et al. 2012). Karyotype evolution is thought to be driven by chromosome fusions and fissions (Karasawa 1979; Cox et al. 1998) although there is some evidence (McQuaid 1949; Lan and Albert 2011) that more complex changes may also play a role. Section *Barbata* is thought to be the most derived section of the genus (Atwood 1984) and many of its species are probably a product of a recent radiation in the island rich region of Malesia. Although studies on natural hybridisation are lacking owing to the rarity of *Paphiopedilum* in the wild, homoploid hybridisation may also play a role in speciation processes as many putative natural hybrids (at least 17 of which involve section *Barbata* members) have been reported and a hybrid origin is suspected for some species and varieties (Cribb 1998).

1.1 Classification of *Paphiopedilum*

The genus *Paphiopedilum* (Cypripediodeae: Orchidaceae) refers to a group of some 80 evergreen tropical and sub-tropical orchids native to Southeast Asia, the Pacific Islands, South China and India (Fig. 1.1). They are closely related to *Phragmipedium*, *Selenipedium* and *Mexipedium* of tropical and sub-tropical Americas as well as *Cypripedium* of temperate North

America, Europe and Asia. These genera are collectively known as lady's slipper orchids after their highly-modified 'kettle trap' flowers.

The major infrageneric classifications of *Paphiopedilum* are summarised in Table 1.1. Examples of the flowers of species in subgenera and sections are shown (Fig. 1.2 and 1.3). All of the classifications are in broad agreement on the subgenus *Parvisepalum* and subgenus *Brachypetalum* as being distinct from the rest of *Paphiopedilum*. The difference is mainly in the treatment of the level and rank of divisions for the remaining members of *Paphiopedilum*. For example Karasawa and Saito (1982) uses subgenus *Sigmatopetalum* to designate the tessellated-leaf tropical *Paphiopedilum* while Cribb (1998) prefers to use section *Barbata* for referring to the same grouping.

The classification proposed by Karasawa and Saito (1982) was based on a combination of taxonomic and chromosomal morphology characters and recognised six subgenera, 12 sections and five subsections. Two of the sections, *Barbata* and *Spathopetalum*, were further divided into subsections. In contrast, the later classification of Cribb (1998), which incorporated taxonomy and an early molecular phylogeny of Cox et al. (1997), based on nucler ribosomal DNA sequences, recognised instead three subgenera with the largest subgenus *Paphiopedilum* being further divided into five distinct sections. This basic classification structure was then modified by Averyanov et al. (2003) with the addition of section *Emersonianum* and section *Parvisepalum* in subgenus *Parvisepalum*. In the same year, Braem and Chiron (2003) published a classification building on that by Karasawa and Saito (1982) but recognising an additional section, adding the rank of 'complex' and also additional species and variety names.

The most recent major update to the classification of *Paphiopedilum* was made by Górniak et al. (2014) for the purpose of accommodating the newly described *P. canhii* (Averyanov et al. 2010) the unusual morphology of which did not fit any of the groups recognised in existing classifications. Utilising a combination of morphology, chromosomal and molecular information from the low-copy nuclear gene *Xdh*, nuclear ITS and plastid markers, they argued that *P. canhii*, which had temporarily been placed in section *Barbata* by Averyanov et al. (2011) was indeed distinct enough to warrant creation of a new subgenus *Megastaminodium* as proposed by Braem and Gruss (2011). As the inclusion of *P. canhii* caused paraphyly in subgenus *Paphiopedilum* (as defined by Cribb 1998) in their *Xdh* and ITS datasets, Górniak et al. (2014) further recommended that the sections within subgenus *Paphiopedilum* be elevated into individual subgenera. For this study, we follow the

classification by Cribb (1998) and also use *Megastaminodium* by Górniak et al. (2014) for the sake of simplicity.

Karasawa and Saito (1982) and Braem and Chiron (2003) have proposed divisions at the subsection and complex level. Subsection- and complex-level divisions for section *Barbata* are presented in Table 1.2 and Table 1.3. Although we consider the use of such low-level divisions to be unnecessary and unwieldy, they are nonetheless useful for giving insight into these workers ideas on inter-taxon relationships within section *Barbata*.

1.2 Distribution, habitat and biology

Subgenus *Parvisepalum* and subgenus *Brachypetalum* are sometimes regarded as the early diverging subgenera (e.g. Cox et al. 1997) and show a continental Asian distribution. The centre of subgenus *Parvisepalum* diversity lies in south-west China and north Vietnam while subgenus *Brachypetalum* appears to be concentrated around south China, Indo-China, Burma and Thailand. Section *Paphiopedilum* ranges from the Himalayas to south China and most of Indochina. The tiny section *Cochlopetalum* is found only in Sumatra and Java while section *Coryopedilum* is found mainly in Borneo. The sections *Pardalopetalum* and *Barbata* are concentrated around Malesia, a biogeographical and phytogeographical region spanning the Malay Peninsula, the Malay Archipelago, Papua New Guinea and the Bismarck Archipelago.

Paphiopedilum have been the subject of intense horticultural interest. Unfortunately, over-collection from wild populations and habitat destruction has resulted in many species becoming rare and threatened. Today, *Paphiopedilum* species generally occur as scattered isolated colonies often consisting of just a small number of individuals (Cribb 1998). However, highly inaccessible areas, such as the *P. niveum* colonies on the limestone islands of the Malaysian Langkawi archipelago, are still known to harbour clumps of several hundred individuals (Chew MY pers. comm.).

The habitat and ecological niches of *Paphiopedilum* are reviewed in detail by Cribb (1998) and Averyanov et al. (2003). Many *Paphiopedilum* species are narrow endemics with many taxa being known only from one or a few locations. Only a handful, such as *P. appletonianum*, *P. bullenianum*, *P. callosum*, *P. concolor*, *P. lowii*, *P. parishii*, *P. philippinense* and *P. villosum* are recorded to have wide distributions. Generally, the plain-leaf groups, consisting of section *Cochlopetalum*, section *Coryopedilum* and section *Paphiopedilum*, occur

in sunny environs while the tessellated-leaf groups of section *Barbata*, subgenus *Brachypetalum* and subgenus *Parvisepalum* prefer more shaded environments.

Section *Barbata* are predominantly wet forest terrestrials that grow under shade on decaying leaves. Only a few taxa, such as *P. hookerae* show broader tolerance for other strata has been reported growing in acidic peat, saline mangrove, ultramafic, serpentine, limestone and sandstone. Subgenus *Parvisepalum*, subgenus *Brachypetalum* and section *Coryopedilum* members are mainly calcicolous lithotrophs, inhabiting the detritus-filled nooks and crannies of limestone karst hills and cliffs. Only a few, such as *P. villosum*, have been encountered epiphytically on trees. *Paphiopedilum* habitats are characterised by high precipitation usually broken by extended dry seasons. Flowering is thought to occur in response to a temperature change i.e. onset of seasonal monsoon rains. While plants in glasshouse culture show marked peak flowering times (Cribb 1998), flowering times are inconsistent in the wild can vary according to local microclimate especially in equatorial tropics where there is no clear cut dry and wet seasons (personal observation). For instance, on the field trip to Borneo, we observed *P. hookerae* plants in flower at the foot of Kinabalu Park but not at lower altitudes at the nearby Kipandi and Poring even though all plants were derived from the same original population.

As a rule, *Paphiopedilum* plants typically produce a single inflorescence bearing a single flower although particularly vigorous plants in culture sometimes produce multiple flowers. Inflorescences with multiple flowers and simultaneous-blooming are only regularly encountered in sections *Coryopedilum* and *Pardalopetalum*, and multiple flowers and successive-blooming in section *Cochlopetalum*. *Paphiopedilum* flowers are long lived and, in the absence of pollination, can last for many weeks.

Deceptive pollination is the rule throughout the genus *Paphiopedilum* (Table 1.4). Subgenus *Parvisepalum*, whose flowers are likened to those of the allied *Cypripedium*, produce sweets scented and bright coloured flowers which mimic the signals of food-rewarding flowers that attract bees. However, *P. armeniacum*, which produces bright yellow flowers, appears to have evolved a generalist pollination mechanism and is visited by several species of bees, primarily halictid bees, as well as hoverflies (Liu et al. 2010).

In the rest of the genus *Paphiopedilum,* hoverfly-pollination predominates. Section *Paphiopedilum* and some members of section *Pardalopetalum* probably produce visual (e.g. shiny yellow staminodes, lurid colours) and olfactory cues that mimic food sources (Fig. 1.4) of hoverflies (Syrphidae) (Bänziger 1996). In section *Coryopedilum* (Atwood 1985), section

Barbata (Bänziger 1996, 2002) and some members of section *Pardalopetalum* (Shi et al. 2007), brood site deception is implicated. Flowers show a combination of visual (e.g. dark red warts) and olfactory signals that mimic the presence of aphid colonies (Fig. 1.5), the natural prey of Syrphidae larvae, and trick female hoverflies into ovipositing on the flower. The larvae that emerge from these eggs invariably die of starvation. It is speculated that there is probably some overlap between food and brood-site deception signals i.e. 'food for you vs. food for your young' (Bänziger et al. 2012).

Most members of subgenus *Brachypetalum* feature putrid-smelling white or cream-coloured flowers covered with dark spots that are pollinated by milesiine hoverflies (Bänziger et al. 2012). This mechanism of attraction here is still not understood but probably differs from above as the larvae of milesiine hoverflies do not feed on aphids. The taxa *P. niveum* and *P. thaianum* show a shift toward bee pollination (Bänziger et al. 2012). These taxa produce faintly fragrant white flowers that are distinct from the rest of subgenus *Brachypetalum* and are pollinated by halictid and meliopine bees respectively.

The pollination process in *Paphiopedilum* is described in detail by Bänziger (1996, 2002). In both food and brood site deception, visiting insects are induced to tumble into the slipper. Once inside, slipper architecture discourages exit from the entrance whilst hairs on the back wall guide the trapped insect through a passage where it must squeeze past the stigma and later one of the two sticky pollinia-bearing anthers to gain exit. Pollination occurs if the pollinium successfully detaches onto the back of insect and the process is repeated in another *Paphiopedilum* flower. The width of the escape passage is closely matched to the size of the pollinator species: 'wrongly' sized insects either exit without affecting pollination or get stuck and perish. *Paphiopedilum* flowers wilt shortly after pollination, leaving behind swollen seed capsules on the inflorescence. After several months, mature capsules dehisce and liberate thousands of fine dust-like seeds into the wind.

Like most orchids, *Paphiopedilum* form intimate species-specific associations with mycorrhiza fungi (Tulasnellaceae) (Athipunyakom et al. 2004; Yuan et al. 2010). They are myco-heterotrophs for much of their early life (Rasmussen 1995): *Paphiopedilum* seeds lack any stored food and only germinate when the seed becomes infected by the fungus. The developing orchid seedling is essentially a parasite and "steals" carbohydrates and other essential nutrients from the fungus to fuel its own growth and development. The orchid-fungus symbiosis later assumes a more mutualistic character as the seedling becomes photosynthetic and begins to transfer carbohydrates to the fungal partner in exchange for minerals. However, a parasitic relationship may persist under certain conditions and is thought

to play a role in enabling terrestrial orchids like *Paphiopedilum* to thrive under light-deprived conditions on the forest floor.

Although studies are lacking, seedling mortality in nature is thought to be high as successful colony establishment is dependent on the combined presence of suitable abiotic factors, pollinators and fungal symbionts (Cribb 1998). However, once established, individuals of *Paphiopedilum* species are long-lived and can potentially persist clonally for decades.

1.3 Cytogenetics

Paphiopedilum possess unusually large chromosomes for orchids (e.g. *P. niveum* has a mean chromosome length of 9.4 μm) (Kamemoto et al. 1963). The basic cytology of *Paphiopedilum* is reasonably well studied and chromosome numbers have been published for many species (Karasawa 1979, 1986; Karasawa and Aoyama 1980, 1988; Karasawa and Tanaka 1980, 1981; Karasawa and Saito 1982; Cox et al. 1998) (see Table 1.5) and indicate a basal diploid karyotype of 26 metacentric chromosomes.

The basal chromosome number is conserved in the subgenera *Parvisepalum* and *Brachypetalum* as well as in the sections *Pardalopetalum*, *Coryopedilum* and most of section *Paphiopedilum* (except for *P. druryi* and *P. spicerianum*) (Karasawa 1979; Karasawa and Tanaka 1981; Cox et al. 1997). It is thought that fusions and fissions to the ancestral karyotypes are responsible for variation in chromosome numbers for species in section *Cochlopetalum* (2*n*=30-36) and section *Barbata* (2*n*=28-42) (Karasawa 1979; Cox et al. 1998).

Some exceptions this trend have been noted. *Paphiopedilum hookerae* of section *Barbata* for instance has 28 chromosomes, two of which are sub-telocentric. Though studies are lacking, it is speculated that the sub-telocentric chromosomes in this species may have possibly been derived from telocentric chromosomes through the addition of terminal chromatin or by a pericentric inversion (Cox et al. 1997).

The reason for the high incidence of chromosomal fission in section *Barbata* and section *Cochlopetalum* is not known. However, it has been postulated that an inherent structural weakness around centromere regions in the large chromosomes of these sections may be responsible (Cox et al. 1998). It has been noted that high numbers of telocentric chromosomes (and hence fusions and fissions) tend to be associated with species that are geographically isolated with narrow endemic distributions (Cox et al. 1997) and those found on islands (Cribb 1998). There is a positive correlation between increasing chromosome numbers and chromosomal recombination. As this in turn can lead to greater genetic diversity, Stebbins (1971) proposed that cytotypes with higher chromosomal numbers may better able to cope

with the harsh selective pressures during the founding of island populations. This form of selection, possibly acting in tandem with genetic drift, may account for the high chromosome numbers in these species.

Karyotypic changes in *Paphiopedilum* have also been investigated by fluorescence *in situ* hybridisation (FISH) using 25S and 5S nuclear ribosomal DNA (rDNA) probes (Lan and Albert 2011). Overall, the number of 25S rDNA and 5S rDNA were found to vary greatly within the genus. The number of 25S and 5S rDNA sites were generally conserved within the subgenus *Parvisepalum* (25S rDNA=2 or 4 sites in diploid metaphase; 5S rDNA=2 sites in diploid metaphase) and subgenus *Brachypetalum* (25S rDNA=2; 5S rDNA=2), section *Cochlopetalum* (25S rDNA=2; 5S rDNA=4) and section *Barbata* (25S rDNA=2, 5S rDNA=2 or 4) although some position changes were noted. Greater variation in numbers and positions of sites was observed in section *Paphiopedilum* (25S rDNA=2, 5S rDNA=2, 4 or 6) and especially in the closely related section *Coryopedilum* (25S rDNA=2, 4, 6 or 9; 5S rDNA= 4 or 6) and section *Pardalopetalum* (25S rDNA=2, 4, 6; 5S rDNA=4).

1.4 Theories on chromosomal changes & speciation

It is known that certain forms of chromosomal rearrangements can contribute to post-zygotic reproductive isolation between related species. When diploid cytotypes with different chromosomal rearrangements interbreed, structural differences between the parental chromosomes can sometimes lead to meiotic irregularities and hence reduced fertility (underdominance) in the hybrids. However, the extent that chromosomal rearrangements play in speciation processes is still unclear and is hotly debated. The prevailing view holds that chromosomal rearrangements become fixed after a species has already diverged in allopatry and thus they only play a role in preventing gene flow when the species comes into secondary contact with parental or sister species (Futuyama and Mayer 1980).

Conversely, one school of thought (White 1978; King 1993) proposes that chromosomal changes in themselves may have a primary role in speciation and that the emergence of new chromosomal rearrangements in a population can eventually lead to the emergence of a new species. However, in addition to the lack of strong supporting evidence, this model of speciation also suffers from the inherent difficulty of explaining how a new chromosomal change can increase in frequency within a population as population genetics theory (Walsh 1982) maintains that the probability of a strongly underdominant trait becoming fixed in a population is remote.

Nevertheless, it can be argued that the theory might be plausible where genetic drift is particularly pronounced such as in plants which are hermaphrodite, capable of self-pollination, multiply vegetatively and periodically exist as fragmented populations (Grant 1981). Also, the chances of a chromosomal rearrangement increasing to fixation may increase if the chromosomal rearrangement confers a selective advantage and if it is favoured by meiotic drive (White 1978; King 1993).

Another school of thought is concerned largely with chromosomal rearrangements such as inversions and their ability to selectively suppress recombination in heterozygotes but not in homozygotes. There are several variations of this theory (Felsenstein 1987; Rieseberg 2001; Noor et al. 2001; Navarro and Barton 2003; Kirkpatrick and Barton 2006) but they have in common a theme that chromosomal inversions play a role in speciation by suppressing recombination and thus capturing and "locking" together combinations of genes that contribute to reproductive isolation.

In any case, more work is needed to link and test these chromosomal speciation theories with real world evidence. A detailed study of the chromosomal rearrangements in *Paphiopedilum* section *Barbata* and its associated hybrids may shed light on the role of karyotypic change in speciation processes.

In the work presented here, I ask the question, 'What is driving speciation in *Paphiopedilum* section *Barbata*?' specifically, in Chapter 2 I ask: (1) what are the evolutionary relationships between the members of section *Barbata*? (2) is hybridisation occurring, and if so, to what extent and which taxa are implicated? I attempt to answer this question by analysing gene trees produced from the sequences of four plastid (*ycf1*, *matK*, *psa-ycf3ex3* and *trnF(GAA)-ndhJ*) and two low copy nuclear gene (*Xdh* and *CHS*) regions.

In Chapter 3, I ask: what biogeographical forces are shaping the diversification of section *Barbata*? To answer this question I estimate the age of *Barbata* from a dated BEAST phylogeny of cloned *Xdh* sequences and analyse it in relation to the geographical history of Southeast Asia.

In Chapter 4, I ask: what genomic changes are occurring in section *Barbata*? I approach this question by characterising the repetitive DNA sequences generated from Illumina HiSeq genome skimming of representative taxa and analysing it against new flow cytometric genome size estimates and published chromosome numbers.

Tables

Table 1.1 Summary of the major infrageneric classifications in *Paphiopedilum*

Karasawa and Saito (1982)*		Cribb (1998)		Averyanov et al. (2003)		Braem and Chiron (2003)*		Gorniak et al. (2014)
Subgenus	**Section**	**Subgenus**	**Section**	**Subgenus**	**Section**	**Subgenus**	**Section**	**Subgenus**
Parvisepalum		*Parvisepalum*		*Parvisepalum*	*Parvisepalum* *Emersonianum*	*Parvisepalum*		Not Applicable
Brachypetalum		*Brachypetalum*		*Brachypetalum*		*Brachypetalum*		Not Applicable
Paphiopedilum	*Paphiopedilum* *Strictopetalum* *Thiopetalum*	*Paphiopedilum*	*Paphiopedilum*	*Paphiopedilum*	*Paphiopedilum*	*Paphiopedilum*	*Paphiopedilum* *Strictopetalum* *Thiopetalum* *Ceratopetalum*	*Paphiopedilum*
Polyantha	*Mastigopetalum* *Mystropetalum* *Polyantha*		*Coryopedilum* *Pardalopetalum*		*Coryopetalum* *Pardalopetalum*	*Polyantha*	*Mastigopetalum* *Mystropetalum* *Polyantha*	*Coryopedilum* *Pardalopetalum*
Cochlopetalum			*Cochlopetalum*		*Cochlopetalum*	*Cochlopetalum*		*Cochlopetalum*
Sigmatopetalum	*Barbata* *Blepharopetalum* *Planipetalum* *Punctuatum* *Sigmatopetalum* *Spathopetalum*		*Barbata*		*Barbata*	*Sigmatopetalum*	*Barbata* *Blepharopetalum* *Planipetalum* *Punctuatum* *Sigmatopetalum* *Spathopetalum*	*Sigmatopetalum*
								Megastaminodium

*subsection and complex level divisions not shown

Table 1.2 Summary of subsectional divisions within section *Barbata*/subgenus *Sigmatopetalum* as proposed by Karasawa and Saito (1982)

Section	Subsection	Species
Barbata	*Barbata*	*P. argus* *P. barbatum* *P. callosum* *P. hennisianum* *P. lawrenceanum*
	Loripetalum	*P. dayanum* *P. ciliolare* *P. superbiens*
	Chloroneura	*P. acmodontum* *P. javanicum* *P. urbanianum* *P. virens*
Blepharopetalum		*P. bouganvilleanum* *P. mastersianum* *P. violascens* *P. wentworthianum*
Planipetalum		*P. purpuratum* *P. sukhakulii*
Punctatum		*P. tonsum*
Sigmatopetalum		*P. venustum*
Spathopetalum	*Macronodium*	*P. hookerae*
	Spathopetalum	*P. appletonianum* *P. bullenianum* *P. celebesense**

*Synonym:
P. celebesense – P. bullenianum var. *celebensense*

Table 1.3 Summary of subsectional and complexes within section *Barbata*/subgenus *Sigmatopetalum* as proposed by Braem and Chiron (2003)

Section	Subsection	Complex	Species
Barbata	*Barbata*		*P. argus* *P. barbatum* *P. crossii**
		lawrenceanum	*P. lawrenceanum* *P. fowliei* *P. hennisianum*
	Chloroneura		*P. acmodontum* *P. javanicum* *P. schoseri* *P. urbanianum*
	Loripetalum		*P. dayanum*
		ciliolare	*P. ciliolare* *P. superbiens*
Blepharopetalum			*P. mastersianum* *P. mohrianum* *P. papuanum*
		violascens	*P. bouganvilleanum* *P. wentworthianum*
Planipetalum			*P. purpuratum* *P. wardii*
		sukhakulii	*P. sukhakulii* *P. dixlerianum**
Punctuatum		*tonsum*	*P. braemii** *P. tonsum*
Sigmatopetalum			*P. venustum*
Spathopetalum	*Macronodium*		*P. sangii*
		hookerae	*P. hookerae* *P. volonteanum**

*synonyms:
P. braemii – P. tonsum
P. crossii – P. callosum
P. dixlerianum – P. × dixlerianum (*P. callosum* × *P. wardii*)
P. volonteanum – P. hookerae

Table 1.4 Summary of known pollinators in *Paphiopedilum*

	Known pollinators	**Reference**
section *Barbata*		
P. callosum	Hoverflies (Syrphidae)	Bänziger 2002
P. purpuratum	Hoverfly (*Ischiodon* sp.)	Liu et al. 2004
subgenus *Brachypetalum*		
P. bellatulum	Milesiine hoverfly (*Eumerus* sp.)	Bänziger 2002
P. concolor	Milesiine hoverfly (*Eumerus* sp.)	Bänziger et al. 2012
P. godefroyae	Milesiine hoverfly (*Eumerus* sp.)	Bänziger et al. 2012
P. niveum	Meliponine bee (*Tetragonula* sp.)	Bänziger et al. 2012
P. thaianum	Halictid bee (*Lasioglossum* sp.)	Bänziger et al. 2012
section *Coryopedilum*		
P. rothschildianum	Hoverfly (*Dideopsis* sp.)	Atwood 1985
section *Pardalopetalum*		
P. dianthum	Hoverfly (*Episyrphus* sp.)	Shi et al. 2007
P. parishii	Hoverfly (*Allographa* sp.)	Bänziger 2002
section *Paphiopedilum*		
P. barbigerum	Hoverflies (*Allograpta* sp. and *Erisyrphus* sp.)	Shi et al. 2009
P. charlesworthii	Hoverflies (Syrphidae)	Bänziger 2002
P. hirsutissimum	Hoverflies (*Allobaccha* sp. and *Episyrphus* sp.	Shi, Luo, Cheng, et al. 2009
P. villosum	Hoverflies (*Betasyrphus* sp., *Episyrphus* sp. and *Syrphus* sp.)	Bänziger 1996
subgenus *Parvisepalum*		
P. armeniacum	Halictid bee (*Lasioglossum* sp.); small carpenter bee (*Ceratina* sp.); hoverfly (*Eristalis* sp)	Liu et al. 2010
P. micranthum	Bumblebees (*Pyrobombus* sp. and *Anthophora* sp.)	Bänziger et al. 2008; Edens-Meier et al. 2015

Table 1.5 Chromosome numbers and distribution of *Paphiopedilum*

Taxa	Chromosome number (2*n*)*	Distribution by botanical country**
section *Barbata*		
P. acmodontum	36 (20)	PHI
P. appletonianum	38 (24)	CHH CHS CBD LAO THA VIE
P. argus	38 (24)	PHI
P. barbatum	38 (24)	THA MLY SUM
P. bougainvilleanum	40 (28)	SOL
P. bullenianum	40 (28), 42 (32)	BOR MLY MOL SUL SUM
P. callosum	32 (12)	CBD LAO MYA THA VIE MLY
P. ciliolare	32 (12)	PHI
P. dayanum	36 (20)	BOR
P. fowliei	36 (20)	PHI
P. hennisianum	36 (18)	PHI
P. hookerae	28	BOR
P. inamoriii		
P. javanicum	38 (24), 40 (28)	BOR JAW LSI SUM
P. lawrenceanum	36 (20)	BOR
P. mastersianum	36 (20)	LSI MOL
P. papuanum		NWG
P. parnatanum		PHI
P. purpuratum	40 (28)	CHC CHH CHS VIE
P. robinsonianum		SUL
P. sangii	28	SUL
P. schoseri	35 (18)	MOL SUL
P. sugiyamanum		BOR
P. sukhakulii	40 (28)	THA
P. superbiens	38 (24), 36 (20)	SUM
P. tonsum	32 (12)	SUM
P. urbanianum	40 (28)	PHI
P. venustum	40 (28)	CHT ASS BAN EHM NEP
P. violascens	38 (24)	BIS NWG
P. wardii	41 (29)	CHC EHM-AP† MYA
P. wentworthianum	40 (28)	SOL
Subgenus *Brachypetalum*		
P. belatulum	26	CHC CHS ASS MYA THA
P. concolor	26	CHC CHS CBD LAO MYA THA VIE
P. godefroyae	26	THA
P. niveum	26	THA MLY
P. thaianum		THA
Section *Cochlopetalum*		
P. glaucophylum	36 (22), 37	JAW
P. moquetteanum		JAW
P. primulinum	32 (14)	SUM
P. victoria-mariae	36	SUM
P. victoria-regina	32, 33, 34, 35, 36	SUM

(cont.)

Table 1.5 Chromosome numbers and distribution of *Paphiopedilum* (cont.)

Taxa	2*n* chromosome number (2*n*)*	Distribution by botanical country**
Section *Coryopedilum*		
P. adductum		PHI
P. gigantifolium		SUL
P. glanduliferum	26	NWG
P. intaniae		SUL
P. kolopakingii	26	BOR
P. ooi		BOR
P. philippinense	26	BOR PHI
P. platyphyllum		BOR
P. randsii	26	PHI
P. rothschildianum	26	BOR
P. sanderianum	26	BOR
P. stonei	26	BOR
P. supardii		BOR
P. wilhelminae		NWG
Subgenus *Megastaminodium*		
P. canhii	26	VIE
section *Paphiopedilum*		
P. areeanum		CHC
P. barbigerum	26	CHC CHS VIE
P. charlesworthii	26	CHC ASS MYA THA
P. coccineum		VIE
P. druryi	30 (8)	IND
P. exul	26	THA
P. fairrieanum	26	ASS EHM
P. gratrixianum	26	CHC LAO VIE
P. guangdongense		CHS
P. helenae		CHS VIE
P. henryanum		CHC CHS VIE
P. hirsutissimum	26	CHC CHS ASS LAO MYA THA VIE
P. insigne	26	CHC ASS BAN?
P. spicerianum	30 (8)	CHC ASS EHM MYA
P. tigrinum	26	CHC MYA
P. tranlienianum		CHC VIE
P. villosum	26	CHC CHS ASS EHM CBD LAO MYA THA VIE
section *Pardalopetalum*		
P. dianthum		CHC CHS VIE
P. haynaldianum	26	PHI
P. lowii	26	BOR JAW MLY SUL SUM
P. parishii	26	CHC ASS LAO MYA THA

(cont.)

Table 1.5 Chromosome numbers and distribution of *Paphiopedilum* (cont.)

Taxa	Chromosome number (2*n*)*	Distribution by botanical country**
Subgenus *Parvisepalum*		
P. armeniacum	26	CHC MYA
P. delenatii	26	CHC CHS VIE
P. emersonii	26	CHC CHS VIE
P. hangianum		CHC VIE
P. jackii		CHC VIE
P. malipoense	26	CHC CHS VIE
P. micranthum	26	CHC CHS VIE
P. vietnamense		VIE

*Compiled from Cox et al. 1997; Kamemoto et al. 1963; Karasawa 1979, 1981, 1986; Karasawa & Aoyama 1988; Karasawa & Saito 1982. Where known, numbers in parentheses indicate number of telocentrics.

**Given as TDWG codes (Brummitt 2001)

Figures

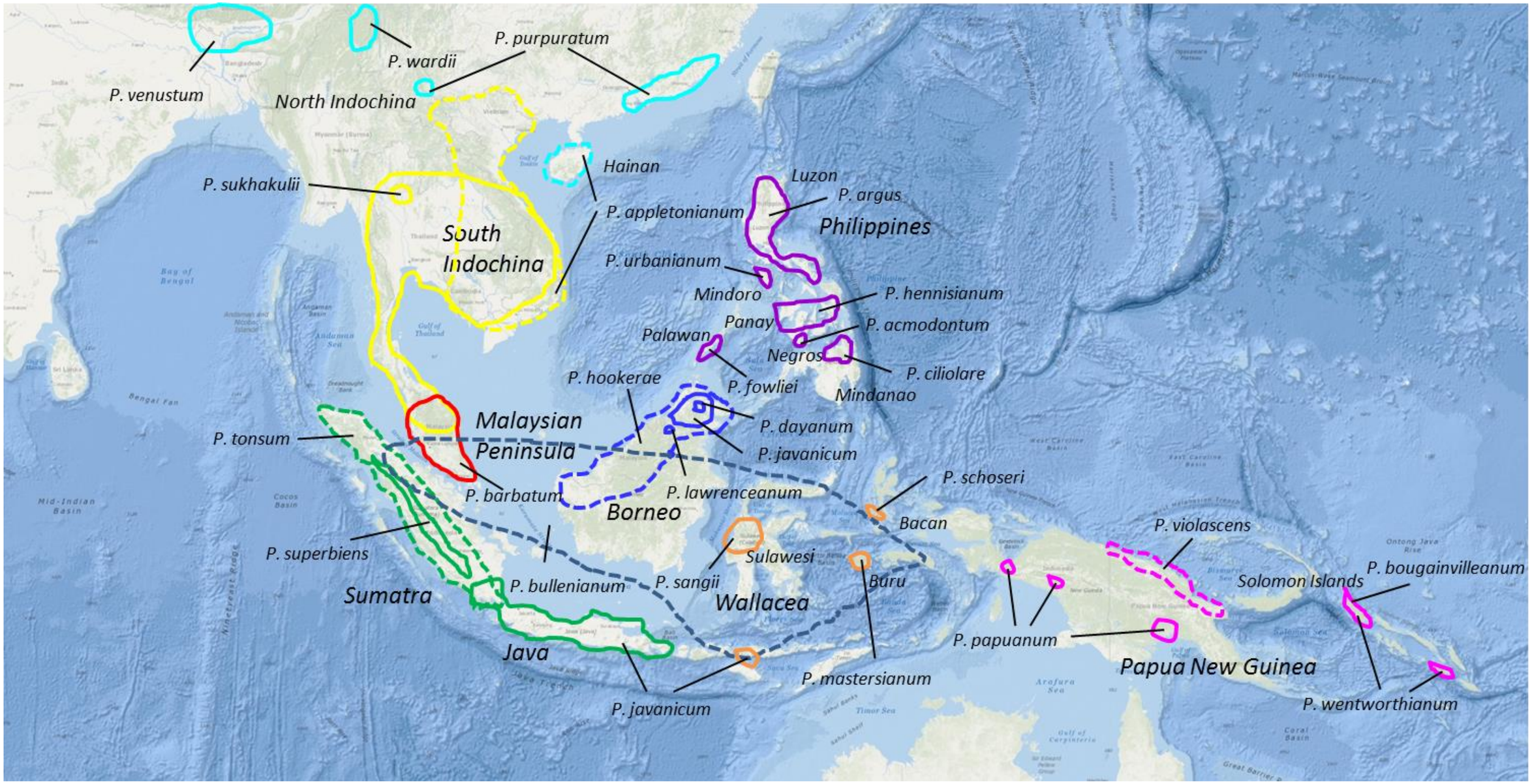

Figure 1.1 Map of SE Asia showing the distribution of *Paphiopedilum* section *Barbata.*

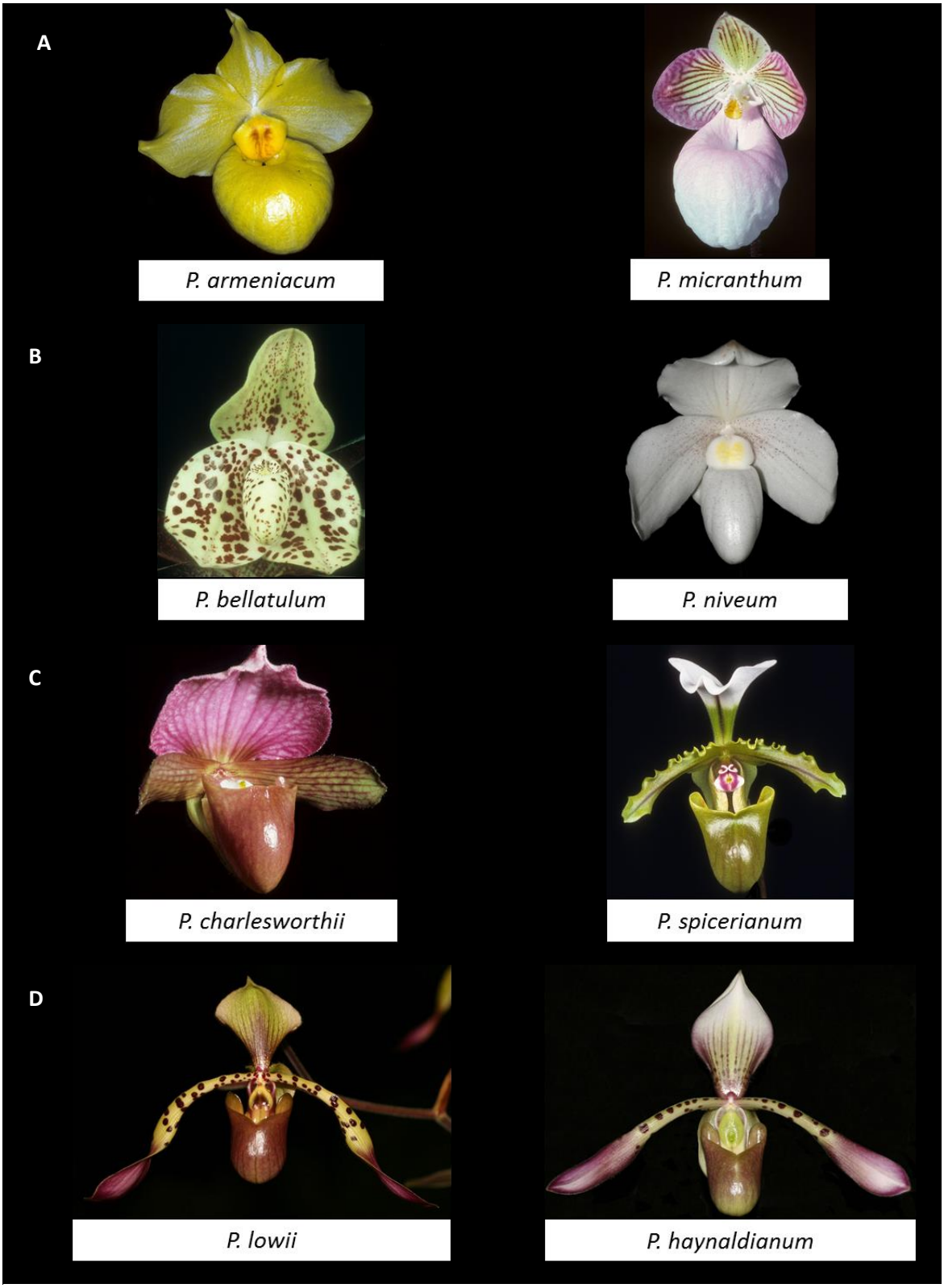

Figure 1.2 Examples of **(A)** subgenus *Parvisepalum*, **(B)** subgenus *Brachypetalum*, **(C)** section *Paphiopedilum*, **(D)** section *Pardalopetalum*, **(E)** section *Cochlopetalum* **(F)** section *Coryopedilum*. Photo credits: *P. armeniacum* (Cribb PJ), *P. micranthum* (Jenny R), *P. bellatulum* (Levy J), *P. niveum* (Ong PT), *P. charlesworthii* (Comber JB), *P. spicerianum* (Levy J), *P. lowii* (Jenny R), *P. haynaldianum* (Jenny R), *P. glaucophylum* (Jenny R), *P. primulinum* (Bryne PO), *P. rothschildianum* (Jenny R), *P. stonei* (Levy J)

(cont.)

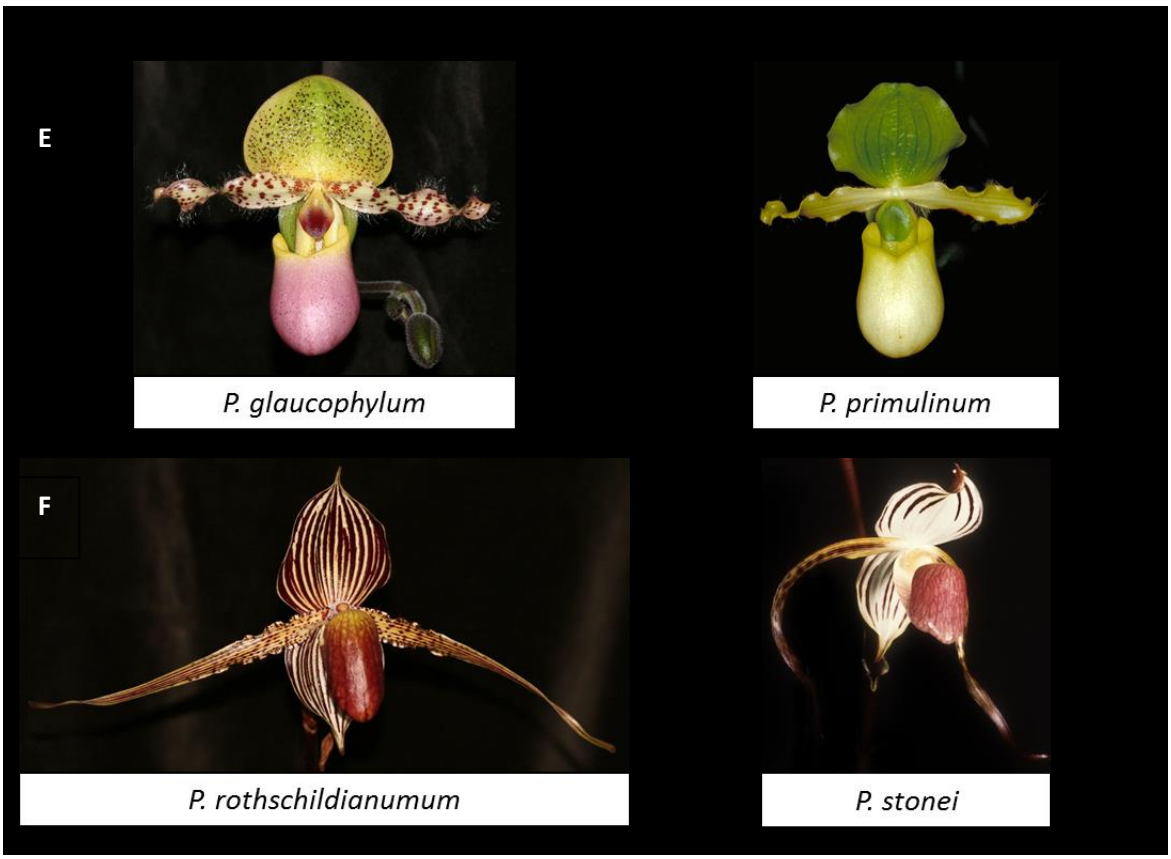

Figure 1.2 (cont.)

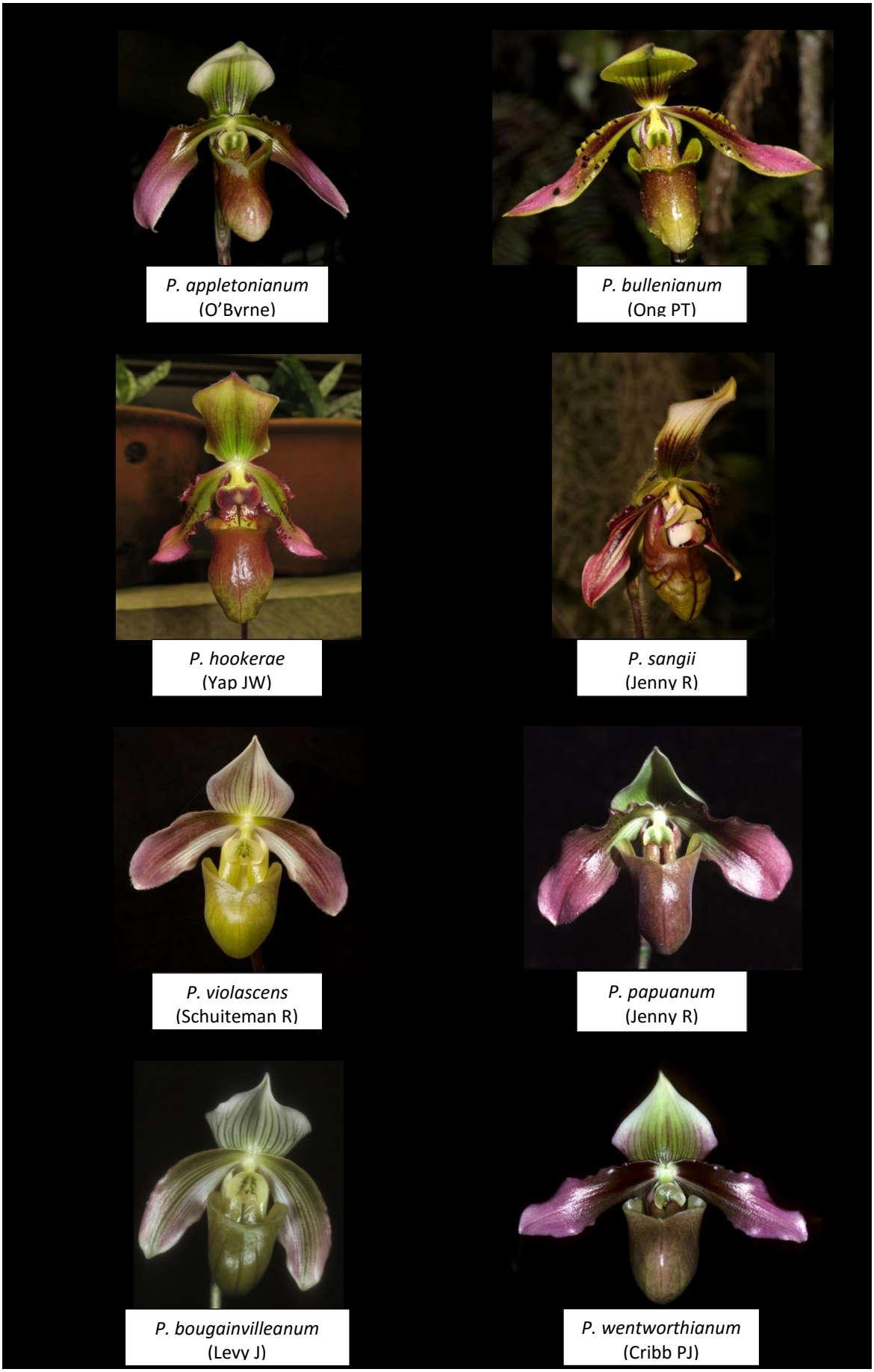

Figure 1.3 Examples of section *Barbata*. Photo credits are in parentheses.

(cont.)

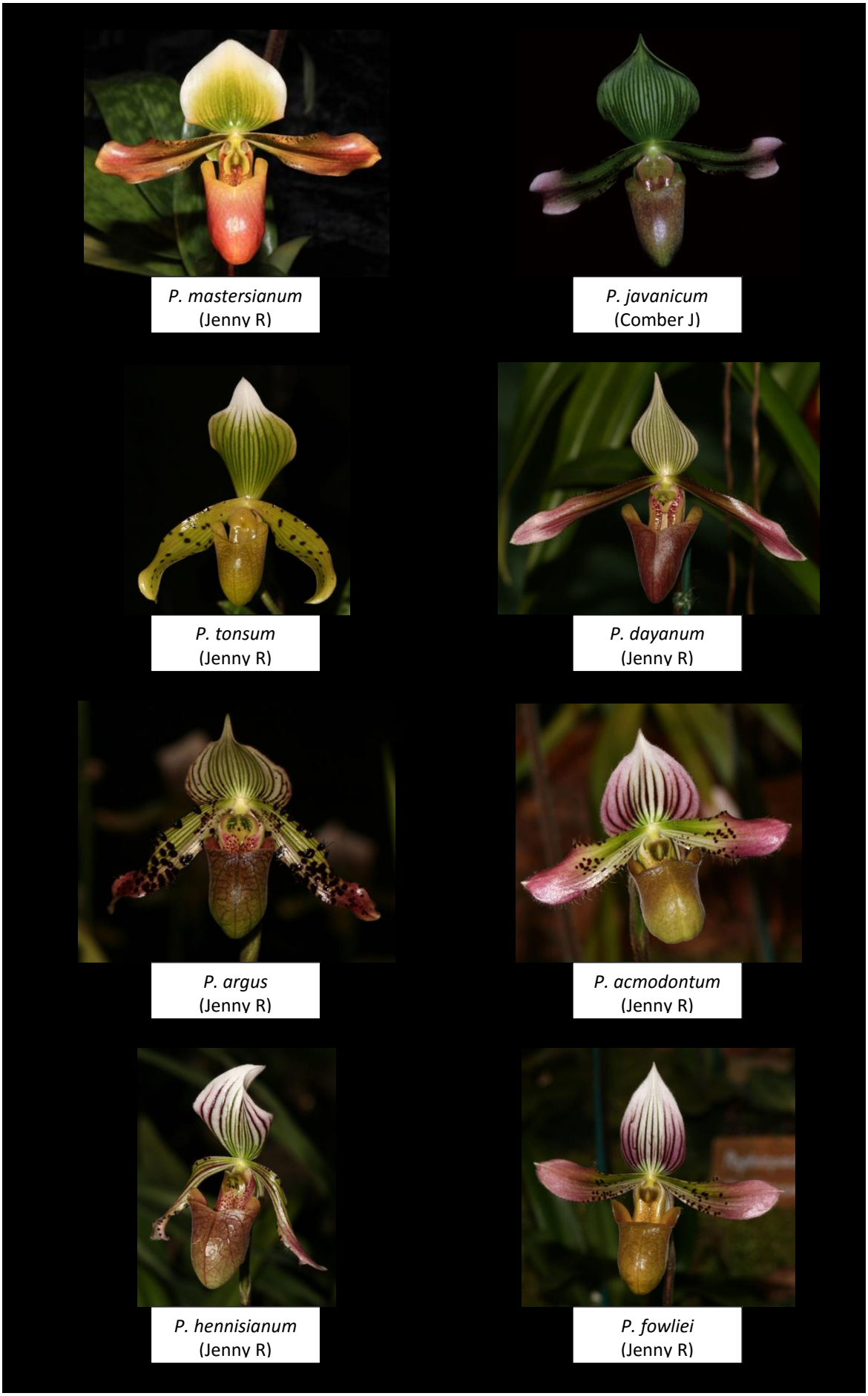

Figure 1.3 (cont.)

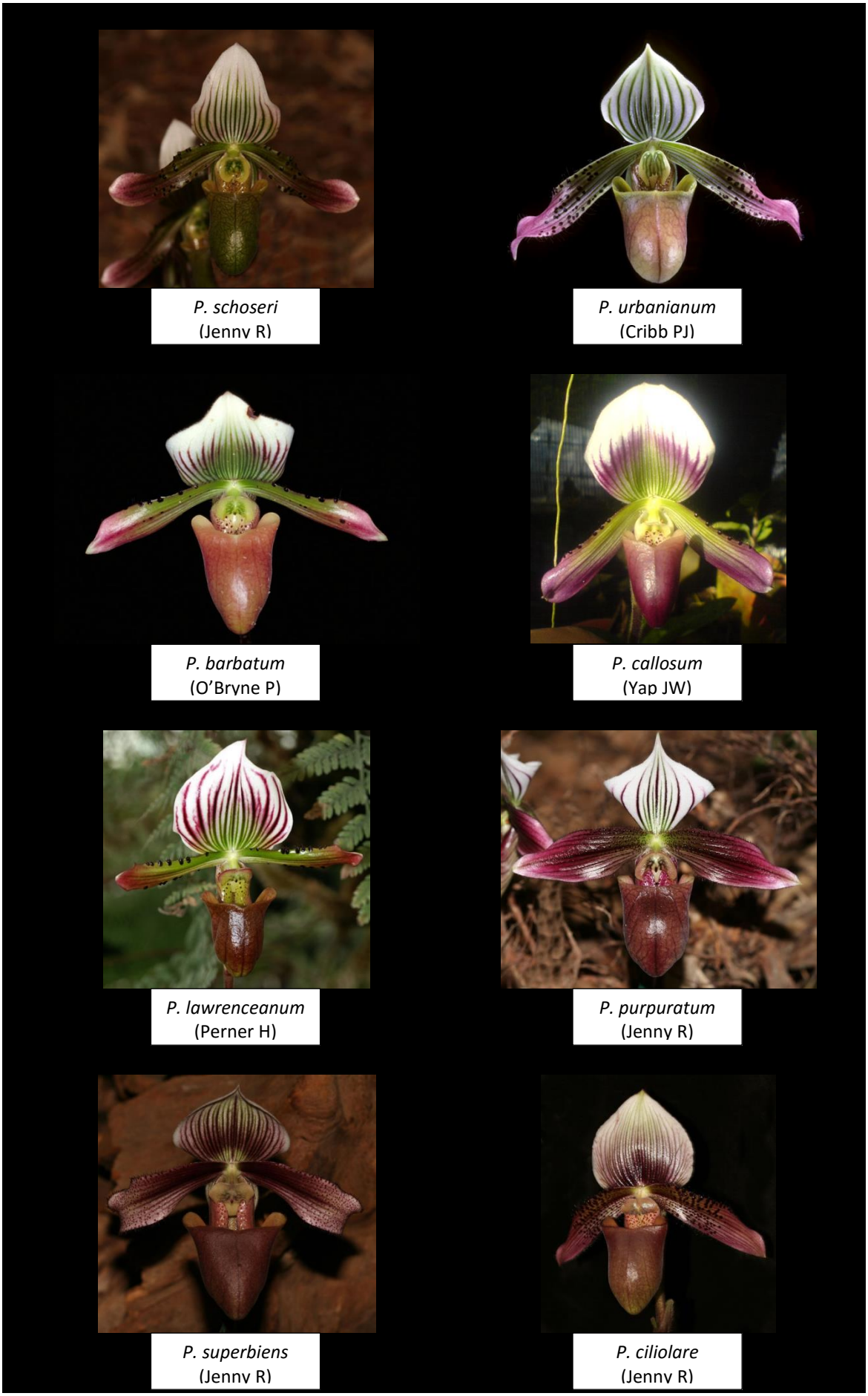

Figure 1.3 (cont.)

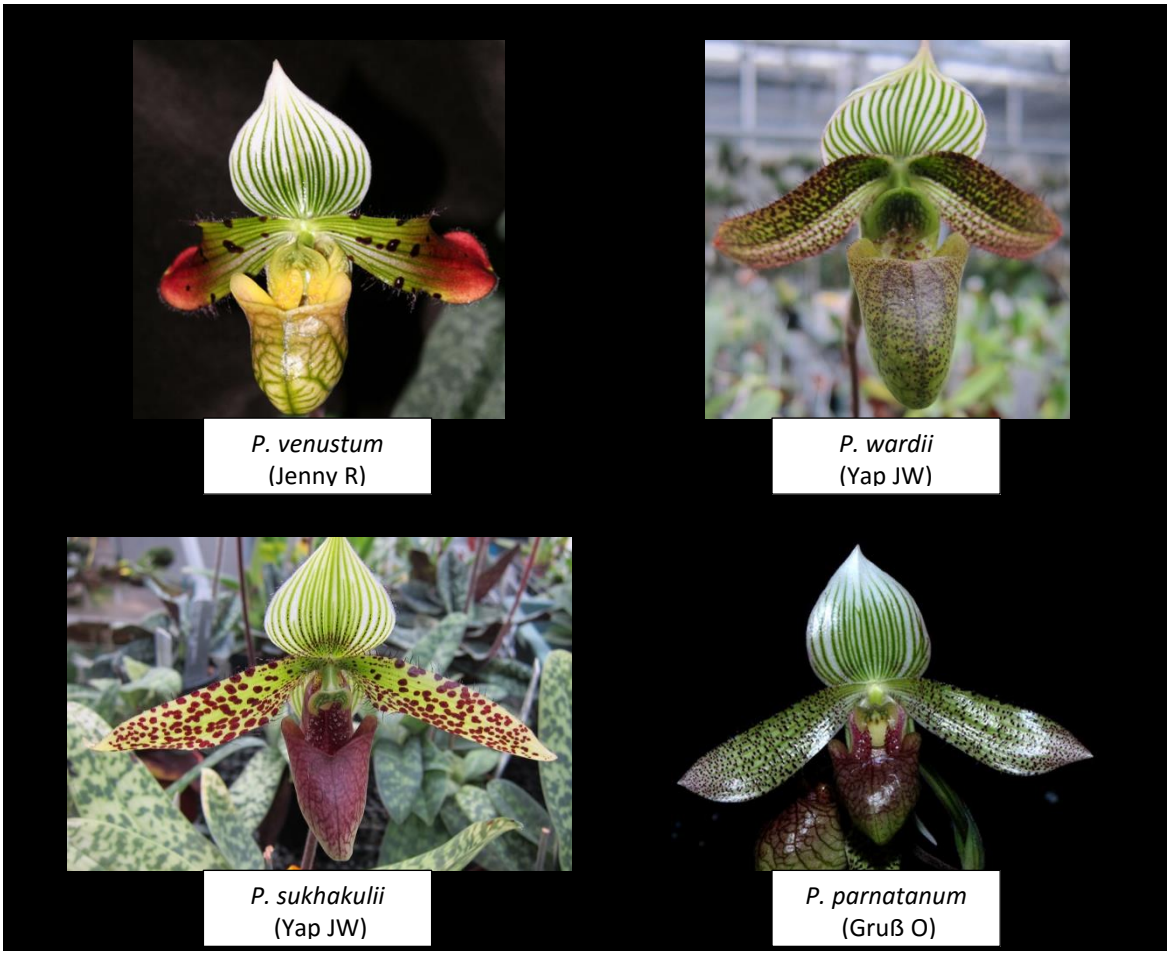

Figure 1.3 (cont.)

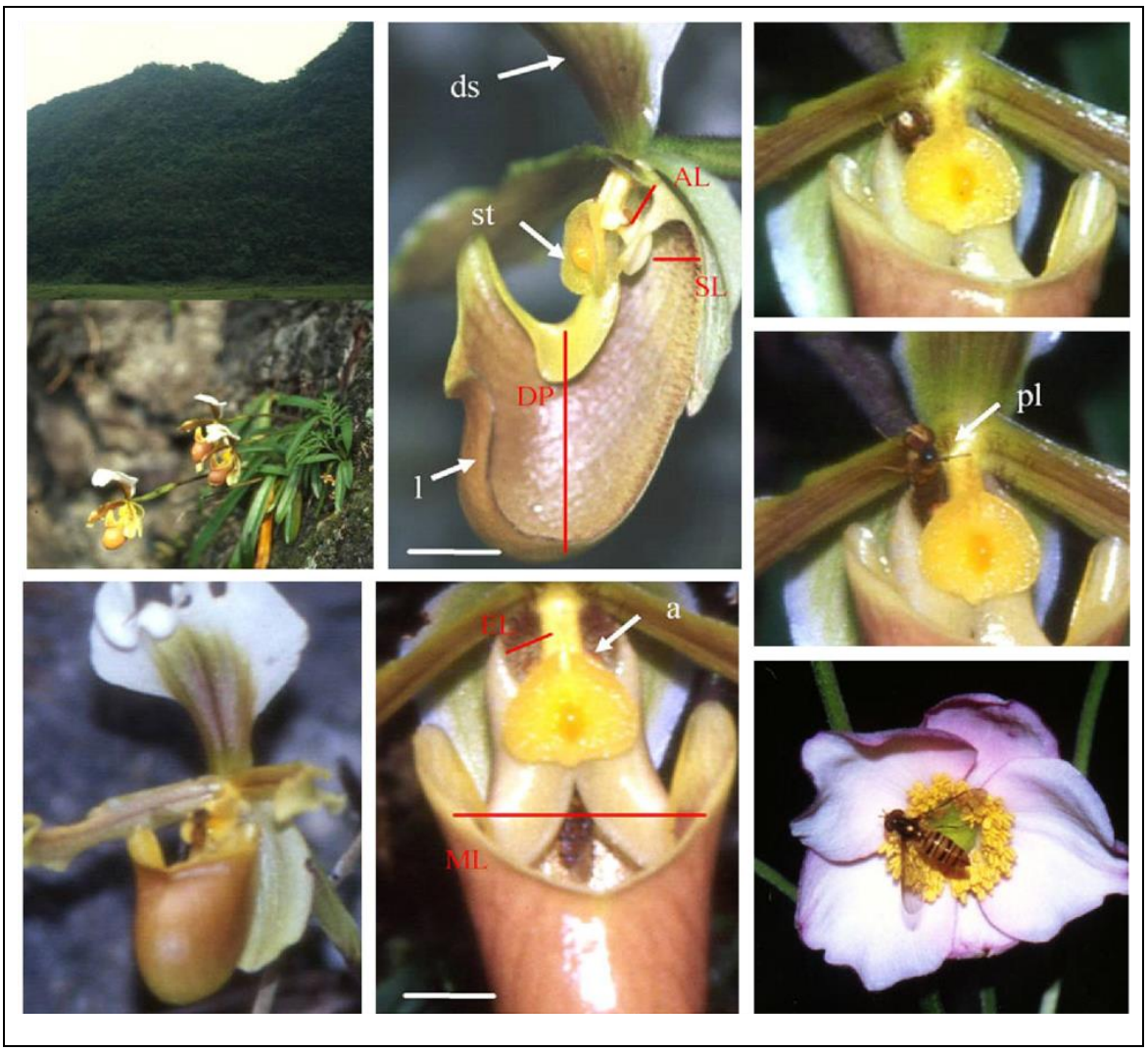

Figure 1.4 Deceptive pollination in *Paphiopedilum barbigerum* of section *Paphiopedilum*. The yellow staminode is thought to attract hoverflies by mimicking food-rewarding flowers. Reproduced from Shi, Luo, Bernhardt, et al. (2009).

Figure 1.5 A flower of *Paphiopedilum schoseri* of section *Barbata*. The red warts are thought to mimic aphid colonies, which are the prey of hoverfly (Syrphidae) larvae. Photo credit: Yap JW

CHAPTER 2
PHYLOGENETICS AND RETICULATE EVOLUTION

2.1 Abstract

P*aphiopedilum* (Cypripediodeae: Orchidaceae) is an early diverging genus of terrestrial orchids for which homoploid hybrid speciation is suspected on the basis of morphological data. We used sequence data from four plastid [*matK; ycf1*; *psaAycf3ex*; *trnF*(*GAA*)-*ndhJ*] regions and the low-copy nuclear genes (*Xdh; CHS*) and sampled multiple individuals per taxon to evaluate the phylogenetic relationships within *Paphiopedilum* section *Barbata*, a biologically diverse and one of the most derived sections of *Paphiopedilum*, to test the hypothesis that hybridisation plays a role in speciation processes in *Paphiopedilum*. I uncovered signals of phylogenetic incongruence providing tentative molecular evidence for historical and on-going hybridization in the evolution of *Paphiopedilum* section *Barbata*. This research has improved phylogenetic resolution between *Barbata* taxa and illustrates the importance of using extensive sampling of multiple individuals and multiple independent assorting loci from both the nuclear and plastid genomes to assess phylogenetic relationships at the species level.

2.2 Introduction

The longstanding horticultural interest in *Paphiopedilum* has given rise to differing opinions on their systematics over the years (Chapter 1, Table 1.1, 1.2, 1.3 and 1.4) with different authors advocating differing levels of subgeneric divsions and interspecific relationships based on morphological, cytogenetic and molecular data. The first comprehensive phylogenetic study of *Paphiopedilum* was by Cox et al. (1997) which assessed the molecular phylogenetics of the subfamily Cypripedioideae using sequences of the internal transcribed spacer (ITS) of nuclear ribosomal DNA (rDNA) by maximum parsimony (MP). This early ITS dataset, which provided the molecular evidence for the infrageneric classification of Cribb (1998) (see Table 1.1), showed good support for the monophyly of subgenera *Parvisepalum*, *Brachypetalum* and sections *Barbata* and *Cochlopetalum*, but only weak support for subgenus *Paphiopedilum*.

Chochai et al. (2012) reassessed the infrageneric relationships within *Paphiopedilum* with ITS and also four plastid regions (*ycf1*, *matK*, *psa-ycf3ex3* and *trnF(GAA)-ndhJ*) with both

MP and newer Bayesian approaches. The study reconfirmed the monophyly of subgenera *Parvisepalum*, *Brachypetalum* and *Paphiopedilum* and sections *Pardalopetalum*, *Paphiopedilum* and *Barbata* across both nuclear and the combined plastid datasets but called into question the monophyly of section *Coryopedilum*, which was supported the Bayesian but not MP, as well as that of section *Cochlopetalum*, which was only supported by ITS but not the plastid loci.

In the same year, Guo et al. (2012) evaluated the biogeography and generic level relationships within subfamily Cypripediodeae using six plastid (*matK*, *rbcL*, *rpoC1*, *rpoC2*, *ycf1* and *ycf2*) and two nuclear loci acyl-CoA oxidase gene (*ACO*), LEAFY (*LFY*) with a combination of MP, Bayesian and maximum likelihood. The analyses demonstrated the subtropical North American genera *Mexipedium* and *Phragmipedium* as sister to the Asian *Paphiopedilum*.

Another phylogenetic study (Górniak et al. 2014), which was spurred by the discovery of *P. canhii* from Vietnam whose morphology defied existing groupings, but was temporarily placed in section Barbata (Averyanov et al. 2010) based on some leaf and flower morphological characters. Górniak et al. (2014) re-evaluated the infrageneric status of *Paphiopedilum* with the inclusion of *P. canhii* using ITS, three plastid (*matK*, *trnH-psbA* and *trnQ-rps16*) and the nuclear gene xanthine dehydrogenase (*Xdh*) loci. The primary result of the study was in demonstrating incongruence between the three loci on the placement of *P. canhii*. This, along with morpological and cytogenetic, was used to justify the creation of subgenus *Megastaminodium* to accommodate *P. canhii*.

The most recent assessment of *Paphiopedilum* was by Guo et al. (2015). The study, which utilised plastid (*accD*, *matK*, *rbcL*, *rpoC2*, *ycf1*, *atpF-atpH*, *atpI-atpH* and *trnS-trnfM*) and the nuclear *ACO*, *LFY*, DEFICIENS-like MADS-box transcription factor (*DEF4*) and DNA repair protein RAD51 (*RAD51*) along with greater taxon sampling, convincingly demonstrated widespread phylogenetic incongruence within *Paphiopedilum* possibly by reticulate evolution. In addition, the study also revealed the existence of polyphyletic species and called into question the monophyletic status of section *Cochlopetalum*.

While the existing studies have greatly clarified infrageneric relationships, species relationships, especially within the largest section *Barbata*, remain unclear due to low taxon sampling and possible misplacement due to reticulation signals. This work extends the findings of existing studies with the objective of further resolving species relationships and clarifying the extent of reticulation through a more comprehensive sampling of *Barbata*.

In this chapter, I ask the following questions: (1) what are the evolutionary relationships between the various subgenera/sections in *Paphiopedilum*? (2) what are the evolutionary relationships between the taxa in section *Barbata*? (3) does natural hybridisation play a role in the evolution of section *Barbata*? If so, to what extent and which taxa are implicated? I attempt to answer these questions by focusing on phylogenetic trees developed using sequences from four plastid regions [*matK*, *ycf1, psa-ycf3ex3* & *trnF(GAA)-ndhJ*] and two low copy nuclear genes *Xdh* and *CHS* (chalcone synthase), which have previously been used in studies of Orchidaceae (Górniak et al. 2010, 2014; Inda et al. 2010). To achieve this I have put significant effort into collecting multiple individuals per taxon in section *Barbata*, with collections sourced from many places (Table 2.1), and analysing them with multiple independent loci.

2.3 Materials & Methods

2.3.1 Taxon sampling

Three overseas collecting trips were made to SE Asia to collect DNA materials from wild plants propagated at (a) the Forest Research Institute Malaysia, Selangor, Malaysia, (b) Kinabalu Park, Sabah, Malaysia, (c) Poring Hot Springs, Sabah, Malaysia, (d) Kipandi Park, Sabah, Malaysia and (e) Bogor Botanic Gardens, Bogor, Indonesia. In addition, samples were collected from cultivated plants propagated in living collections at (a) the Royal Botanic Gardens, Kew (RBG, Kew), UK, (b) the School of Biological and Chemical Sciences, Queen Mary University of London, UK, (c) Munich Botanic Gardens, Germany, (d) Herrenhausen Gardens, Germany, (e) Ratcliffe Orchids Ltd. UK, and (f) the British *Paphiopedilum* Society. Additional DNA samples were obtained from the DNA Bank, RBG, Kew (http://apps.kew.org/dnabank/homepage.html). As natural interspecific gene-flow is suspected in section *Barbata*, multiple individuals were sampled from each taxon to increase the chances of detecting interspecific hybridisation. Sampling was based on the species and varieties described in scientific and popular literature. As *Paphiopedilum* section *Barbata* has previously been shown to be monophyletic and the *Phragmipedium* as a sister to the genus *Paphiopedilum* (Cox et al. 1997; Chochai et al. 2012; Guo et al. 2012), *Phragmipedium besseae* and members of other *Paphiopedilum* sections were used as out-groups. I obtained a total of 139 samples covering 108 *Paphiopedilum* section *Barbata* and 31 out-group representatives. A list of taxa analysed and their geographical origins is given in Table 2.1.

2.3.2 Primer design

The new primers in Table 2.2) were designed on Primer3 (Untergasser et al. 2012) on Geneious v.7.1.8 using sequences from a draft transcriptome of *P. primulinum* (unpublished data provided by Dr. Yung-I Lee) as well as from sequences published on public domains. In addition to *Xdh* and *CHS*, new primers were developed for amplifying CYCLOIDEA (results not shown). However, this region was excluded from analyses because the product generated by PCR were too short to be useful (<>300 bp). Primers against ITS, used in previous analyses, generated highly heterogeneous sequence populations for species in section *Barbata*, indicative of incomplete homogenisation of units and the coexistence of ancient and new ITS variants in individual species. For this reason these sequences were also excluded from analyses.

2.3.3 DNA extraction, PCR and sequencing

Sample DNA was extracted from fresh or desiccated leaves using a 2× cetyl trimethylamonium bromide (CTAB) method developed by Doyle & Doyle (1987). Extracts were purified by either caesium chloride/ethidium bromide density gradients or DNA purification columns (NucleoSpin Extract II Columns; Macherey-Nagel, GmbH & Co. KG, Germany) following the manufacturer's recommendations. The DNA regions analysed in this study include: (1) approximately 2300-bp combined total of two fully coding plastid gene regions *ycf1* and *matK*, (2) approximately 1100-bp combined total of two non-coding plastid regions *psa-ycf3ex3* and *trnF*(GAA)-*ndhJ*, (3) approximately 1005-bp of a fully coding region of the low-copy nuclear gene *Xdh*, and (4) approximately 736-bp of a fully coding region of the low-copy nuclear gene *CHS*. Details of primers and PCR mixes and conditions are given in Tables 2.2, 2.3 and 2.4. PCR products were cleaned with DNA purification columns and subjected to cycle-sequencing in both directions using a Big Dye Terminator Kit (Applied Biosystems Inc, Warrington, UK) according to the manufacturer's protocol. Cycle-sequencing products were cleaned by ethanol precipitation and loaded into an ABI 3730 automated sequencer. Raw sequence reads were edited, assembled and aligned into individual matrices according to DNA region using Geneious v7.1.8 (Biomatters Ltd, Auckland, NZ)(Kearse et al. 2012) software. All sequences were inspected visually for accuracy.

2.3.4 Cloning

Xdh samples which displayed evidence of multiple copies in electropherogram peaks, were selected for cloning and sequencing. The *Xdh* DNA template for cloning was prepared in two-steps. Genomic DNA was first amplified using PCR with Phusion® high fidelity DNA polymerase

and Xp551F and Xp1590R primers (Table 2.2). A 1/10 dilution of the initial PCR product was then used as the template and amplified with nested PCR primers. Details of PCR conditions and the reaction mixes used are given in Tables 2.2 and 2.3 respectively. PCR products were cleaned using DNA purification columns and checked for quality and concentration using gel electrophoresis and spectrophotometry.

PCR products were cloned into pGEM T-easy (Promega) following the manufacturer's instructions using a ligation mix consisting of 5 µL 2x ligase buffer, 1 µL pBluescript SKII (Agilent Technologies 1 µL T4 DNA ligase and 3 µL PCR product. The mix was incubated at 4°C for 24 h. Chemically competent DH5α^{TM} cells were transformed with 5 µL of ligation reaction mix. Cells were heat shocked for 50 s at 42˚C, kept on ice for 2 min and incubated with 500 µL of LB medium at 37˚C with gentle agitation. The suspension was plated on to LB agar plates containing 0.5 mM IPTG, 80 µg/mL X-gal and 100 µg/ml ampicillin and incubated at 37˚C overnight.

White (transformed) colonies were selected, resuspended in LB-ampicillin and incubated overnight at 37°C with gentle agitation. The resulting bacterium cultures were pelleted and the plasmid DNA purified (using a plasmid prep kit Thermofisher Scientific). Plasmid DNA was then digested with EcoRI (NEB) for 2 h at 37°C, separated on a 1.5% agarose gel containing 5 µg/mL ethidium bromide and visualized under a G:BOX EF (Syngene) in comparison with a ladder (2-log, NEB). Products were sequenced using T7 or SP6 primers (Table 2.2). This process was repeated to give a total of ten clones per *Xdh* sample.

2.3.5 Phylogenetic analyses

Sequence data were arranged into individual matrices according to DNA region. Unreliable reads at sequence ends were trimmed and gaps were treated as missing data. The combined plastid dataset was partitioned into coding *ycf1*, *matK*, *psa-ycf3ex3* and *trnF(GAA)-ndhJ* regions. In the *CHS* dataset, paralogues, as evidenced by multiple electropherogram peak signals at certain base sites, were treated as intra-taxon polymorphisms for the sake of phylogenetic analysis. The cloned nuclear *Xdh* and *CHS* matrices were also screened with RDP v.4.27 (Martin et al. 2015) using default settings to test for possible recombination. Phylogenetic trees were constructed from (1) combined plastid, (2) cloned nuclear *Xdh*, and (3) nuclear *CHS* sequence data by maximum parsimony (MP) and Bayesian methods. A summary of each dataset is shown in Table 2.5. Nodes from all trees were inspected visually for signs of incongruence as evidence for reticulation.

In order to maximise the phylogenetic information in the individual loci, I prepared a Bayesian tree from a partitioned matrix comprised of concatenated plastid and cloned nuclear *Xdh* and *CHS* sequences. The positions of individual sample sequences were examined by eye, before and after concatenation. Samples which showed (1) obvious topological incongruence, (2) a reduction in posterior probabilities (PP) support after concatenation, and (3) introgression, based on strong morphological data, were eliminated from the final concatenated dataset (Table 2.6).

MP analyses were conducted in PAUP* version 4.0b10 for Macintosh (Swofford 2002). All characters were treated as unordered and equally weighted (Fitch 1971). Heuristic searches were conducted with equal character weighting, DELTRAN character-state optimization, performing 1,000 random addition sequence replicates with tree-bisection–reconnection branch swapping, holding ten trees at each step, and MulTrees in effect. Node support was assessed by 2000 bootstrap (Felsenstein 1985) pseudoreplicates. Bootstrap settings were identical to the MP search but with single random addition sequence replicate per pseudoreplicate.

Bayesian analyses were conducted in MrBayes 3.2.2 on CIPRES Science Gateway (Miller et al. 2010) (https://www.phylo.org/portal2). The best-fit model of evolution for each dataset and partition was selected using the BIC score in MrModeltest v2.2 (Nylander 2004). In each search, multiple runs of three Markov Monte Carlo (MCMC) chains (Yang and Rannala 1997) were run sampling every 1000 generations. Each parameter estimation was checked with Tracer v1.4 to confirm effective sampling sizes had been achieved for each parameter and runs had converged.

2.4 Results

2.4.1 Subgenera/section relationships

I used RDP v4.27 (Martin et al. 2015)to screen for possible recombination within nuclear gene regions which can result in anomalous sequences. However, as no recombination events were detected by RDP, all *Xdh* and *CHS* sequences were included in the phylogenetic analysis. Paralogues, as evidenced by sequence ambiguities in the electropherograms, were detected in *Xdh* and *CHS*. Cloning of *Xdh* revealed the presence of up to least eight copies of *Xdh* (Table 3.1). The presence of paralogues might be caused by hybridisation, gene duplication and/or polyploidy, although there is no evidence for recent polyploidy in *Paphiopedilum*.

MP and Bayesian analysis of the sequences produced plastid (Fig. 2.1A-B), cloned *Xdh* (Fig. 2.2A-B), *CHS* (Fig. 2.3A-B) and phylogenetic trees. With some exceptions, e.g. section *Cochlopetalum* and *P. fairrieanum*, there was generally good support for the established *Paphiopedilum* groupings of Cribb (1998). However, apart from section *Parvisepalum*, which was sister to the rest of *Paphiopedilum* in plastid, *Xdh* and *CHS* trees, the relative positions of the other groups were found to differ between loci. A close relationship between multiflowered sections *Cochlopetalum*, *Coryopedilum* and *Pardalopetalum* is supported by the plastid (Fig. 2.1 A-B) and *Xdh* (Fig. 2.2 A-B) and *CHS* trees (Fig. 2.3 A-B). However, there is no support for the monophyly of section *Cochlopetalum* in the plastid, *Xdh* and *CHS* trees. For example, in the *Xdh* trees (Fig. 2.2 A-B), *P. glaucophylum* and section *Paphiopedilum* are sister to a clade comprised of sections *Coryopedilum*, *Pardalopetalum* and the rest of *Cochlopetalum*. These observations agree with findings from Chochai et al. (2012) and Guo et al. (2015).

The taxa *P. canhii* and *P. rungiyasanum* from Vietnam, and *P. fairrieanum* from Bhutan and Sikkim (Fig. 2.8) have unusual morphologies that fall outside the established section/subgenera groupings. In addition, the placement of *P. canhii* and *P. rungiyasanum* is of considerable horticultural and scientific interest because they are newly described (within the last ten years). Unfortunately, we were unable to include *P. rungiyasanum* in our analysis due to difficulties in obtaining legal DNA samples. Previously, *P. fairrieanum* was included in section *Paphiopedilum* based on ITS sequence data (Cox et al. 1997) while *P. canhii* was placed into its own group, subgenus *Megastaminodium* by Górniak et al. (2014) based on a combination of phylogenetic (*Xdh*, ITS and plastid loci), cytological and morphological data. In our study, the plastid trees (Fig 2.1) suggest a close relationship between *P. canhii* to section *Cochlopetalum*, *Coryopedilum* and section *Pardalopetalum,* and place *P. fairrieanum* as sister to a clade containing section *Barbata* and section *Paphiopedilum*. On the other hand, the *Xdh* trees (Fig.

2.2) places *P. canhii* as sister to *P. fairrieanum* which, is in turn, sister to a clade comprised of the sections *Cochlopetalum*, *Coryopedilum* and *Paphiopedilum*, whereas on the *CHS* trees analysed by MB (Fig. 2.3 A), *P. fairrieanum* is placed as sister to *Barbata* with modest support (PP=0.73). Overall our data support Górniak et al. (2014)'s assessment of *P. canhii* as belonging to its own group *Megastaminodium* but also suggest that *P. fairrieanum* may warrant similar recognition as a distinct taxon that does not belong with other established groups. More individuals will need to be analysed to confirm the precise phylogenetic positions of these taxa.

The monophyletic status of section *Barbata* was supported by the plastid (BP=80, PP=1) (Fig. 2.1A-B), *Xdh* (BP=98, PP=1) (Fig. 2.2A-B) and *CHS* (BP 88, PP=1) (Fig. 2.3A-B) analyses. The taxon relationships within the section is expanded upon below.

2.4.2 Relationships within section *Barbata*

A map showing the major geographical distribution of section *Barbata* taxa is shown in Fig. 2.4. Clades representing morphologically (see Table 1.2) and geographically allied taxa were recovered in all phylogenetic trees. However, the clade compositions varied between each dataset and the positions of many taxa remained unresolved on polytomies.

Comparisons of tree nodes between the plastid, *Xdh* and *CHS* trees revealed polyphyly and incongruence in the placement of several taxa and even whole clades (Fig. 2.5, 2.6 and 2.7). There was also incongruence within the cloned *Xdh* dataset between paralogues originating from the same sample (Table 3.1). Incongruence between independent assorting trees and between paralogues may be caused by hybridisation or incomplete lineage sorting. Nevertheless, phylogenetic analyses did recover several distinct and natural clades within section *Barbata*.

Two morphologically similar taxa with the lowest chromosome numbers, *P. hookerae* (2*n*=28) from Borneo and *P. sangii* (2*n*=28) from neighbouring Sulawesi, formed a clade (Clade A) with modest to high support in the plastid, *Xdh*, *CHS* and concatenated trees (Fig. 2.5A-B, 2.6A-B, 2.7B and 2.8) and is sister to all other *Barbata* in the *Xdh* and concatenated trees (Fig. 2.6A-B and 2.8). Another strongly supported clade (Clade B) of distinct taxa with high chromosome numbers, *P. venustum* (2*n*=40) from the Himalayas, *P. wardii* (2*n*=41) from Burma and *P. sukhakulii* (2*n*=40) from South Indochina (Thailand), were found on the plastid and concatenated trees (Fig. 2.5A-B). Good support (PP=0.86) was obtained in the plastid trees to support a clade (Clade C) composed of morphologically allied *P. mastersianum* (2*n*=36) from

Wallacea (Moluccas islands) and *P. violascens* (2*n*=38) and *P. papuanum* (2*n*=?) from Papua New Guinea at the south-western extreme of Malesia. Most samples of the allied *P. appletonianum* and *P. bullenianum* (Clade D) were unplaced. However, clades comprised of *P. appletonianum* and *P. bullenianum Xdh* sequences were recovered (Fig. 2.6A-B). In addition, there was also support for a complex, relationship between the morphologically allied *P. callosum* (2*n*=32) from South Indochina (Thailand), *P. barbatum* (2*n*=38) from the Malaysian Peninsula and *P. lawrenceanum* (2*n*=36) from Borneo. Generally samples of *P. barbatum* and *P. callosum* were found to resolve onto distinct clades (Clade D). However gene tree conflict between the plastid, *Xdh* and *CHS* (Fig. 2.5A-B, 2.6A-B and 2.7A-B) loci suggesting that *P. lawrenceanum* may be hybrid of *P. barbatum* and *P. callosum*.

Among widespread taxa, considerable effort was expended to sample multiple individuals. The outcome was to uncover an interesting phenomenon whereby members of the same taxon were found resolving into clades with morphologically distinct but sympatric taxa (syngameons) rather than with other members of the same taxon. There are several examples. (1) Perhaps the clearest example can be seen in widespread taxa such as *P. appletonianum* (2*n*=38) from North and South Indochina and *P. bullenianum* (2*n*=40) from the Malaysian Peninsula, Borneo, Sumatra, Sulawesi and Ceram. The samples *P. appletonianum* **(**coded A, B C & D) from South Indochina were found to resolve on a clade with *P. callosum* samples from South Indochina (Fig. 2.5). (2) Similarly, *P. bullenianum* H sample from the Malaysian Peninsula was found to group together with *P. barbatum* samples from the Malaysian Peninsula (Fig. 2.6) whereas *P. bullenianum* A from the Malaysian Peninsula and *P. bullenianum* F from Sumatra resolved with *P. superbiens*, *P. javanicum* and *P. tonsum* samples from Sumatra and Java (Fig. 2.5).

In addition, many samples appear to have conflicting phylogenetic signals from different loci i.e. different positions on different trees. For example, samples *P. lawrenceanum* A, B and C from Borneo resolved together with *P. callosum* samples from South Indochina and the Malaysian Peninsula on the plastid tree (Fig. 2.5 A-B). However, on the *Xdh* tree (Fig. 2.6A-B), the same samples were found to resolve on a clade with *P. barbatum* samples from the Malaysian Peninsula.

A tree (Fig. 2.8) was produced from the combination of congruent plastid, nuclear *Xdh* and *CHS* sequences in an attempt to visualise the overall phylogenetic signal on taxon relationships without the effect of reticulation. The combined phylogenetic signals resolve Clade A from Borneo and Sulawesi, and subsequently Clade B, found in Indochina, and then the widespread *P.bullenianum*, as sister to the rest of *Paphiopedilum* section *Barbata* with

good (>0.7) support. Clade C, from Papua New Guinea and Wallacea, was found to have a close phylogenetic affinity to taxa from the Philippines and Borneo.

2.5 Discussion

2.5.1 Relationships between subgenera/sections

Several previous studies have assessed the phylogenetics of *Paphiopedilum* (Cox *et al*. 1997; Chochai *et al*. 2012; Gorniak *et al*. 2014; Guo *et al*. 2015). There is unequivocal support for the monophyly of subgenera *Parvisepalum* and *Brachypetalum*, which is supported in all plastid trees (Chochai *et al*. 2012; Gorniak *et al*. 2014; Guo *et al*. 2015) and in the plastid analyses presented here, and in the ACO, RAD51, DEF4 and *LFY* trees shown by Guo *et al*. (2015).

However there is no consensus for monophyly in subgenus *Paphiopedilum*, with conflicting patterns in the ITS sequences (Chochai et al. 2012; Górniak et al. 2014) and in the *Xdh* and *CHS* sequences (Fig. 2.6 and 2.7). There is also paraphyly in the placement of species in *Cochlopetalum* in all three trees presented here (Fig. 2.1, 2.2 and 2.3). Potentially these conflicts are generated by insufficient sequencing (especially in the case of *CHS* sequences which are 736-bp in length with little polymorphism), but it may also arise through reticulation occurring via intersectional hybridisation or incomplete lineage sorting, (see also section 2.5.2, below).

Cox et al. (1997) postulated that changes in chromosome number in *Paphiopedilum* could be a mechanism for driving diversification in the genus. Differences in chromosome numbers are thought to present a barrier for hybridisation (Robinson 1995), a hypothesis that has received recent support from work on sunflowers (Strasburg et al. 2009), but the strength of the mechanism as a driver of speciation is controversial (Butlin 1993). Certainly signatures of hybridisation were found even between species with differing chromosome number complements such as between *P. callosum* (2*n*=32) and *P. appletonianum* (2*n*=38). Nevertheless, the consistent resolution of the well-supported clade comprising *P. hookerae* and *P. sangii*, both with 2*n*=28, might suggest that certain karyotypes may have a role in preserving the integrity of the clade.

Pollinator shifts are also thought to play an important role in driving orchid diversification (Givnish et al. 2015). Within *Paphiopedilum*, pollinator shifts were documented in subgenus *Parvisepalum* (Liu et al. 2006) and subgenus *Brachypetalum* (Bänziger et al. 2012). However, the present data (Bänziger 1994; 2002) suggest that *Barbata* pollination is via brood-site deception on female hoverflies (Syrphidae). Although conclusive studies are lacking, this pollination strategy is thought to favour outcrossing but without any taxa specificity i.e. there

is nothing to suggest that the pollinator species would prefer to visit only flowers of a member of the same taxon of section *Barbata* over that of another. This lack of pollinator-taxon specificity may help explain the plethora of reticulation signals in our dataset. More work on the brooding behaviour of Syrphidae flies is needed e.g. ranges/distances travelled during brooding.

2.5.2 Relationships within section *Barbata*

This chapter presents the most comprehensive taxon sampling for section *Barbata* to date and includes up to 108 samples from diverse sources (Table 2.1) including multiple samples per taxon. Previously, the largest number of samples analysed were presented in Guo *et al*. (2015, 28 taxa) and Chochai *et al*. (2012, 10 taxa). This improved sampling means that we are able to more confidently predict phylogenetic groups, interpret patterns of reticulation and, in Chapter 3, use the data to reconstruct the biogeographic history of the section.

The overall low phylogenetic depth and reticulation in signals obtained in this study suggest that section *Barbata* is an evolutionarily young group. Thus diversification is probably ongoing and biological barriers to hybridisation are weak. Collectively, the conflicting placements and geographical groupings of species in section *Barbata* suggest that introgressive hybridisation is occurring. Although incomplete lineage sorting may also give rise to tree incongruence and cannot be discounted, the additional supporting biogeographical data and morphology indicate that hybridisation is more likely to account for the phylogenetic patterns observed here. The biogeographical history of the region may help to explain several counter-intuitive groupings, such as the grouping of several *P. bullenianum* samples (G-K) from the Malaysian Peninsula and Sulawesi with *P. sukhakulii* and *P. venustum* samples from north and south Indochina on the *CHS* tree (Fig. 2.7). Modern day Southeast Asia consists of several archipelagos of islands on a shallow continental shelf. However, during each glacial maximum, these became reconnected to each other and to the mainland as a result of receding sea levels. This in turn probably resulted in dramatic range shifts, expansions and contractions. These could have resulted in secondary contact between species and interspecific hybridisation between newly sympatric taxa. Thus, the seemingly unusual phylogenetic signals may be genetic footprints of ancient hybridisation events that hint at past colonisation routes.

Although the fine-powdery seeds in all species of Cypripedioideae are theoretically capable of long-range dispersal (Arditti & Ghani 2000), a recent study suggests that biogeographical vicariance appears to be more likely to account for their modern day distributions (Guo *et al*. 2012). The finding here that *P. lawrenceanum* from Borneo may be a

putative hybrid between *P. barbatum* from the Malaysian Peninsula and *P. callosum* from Thailand in Borneo hints at the influence of past biogeographical events i.e. the repeated connection and isolation of land masses in Malesia, in shaping the diversity in *Barbata*. This biogeographic hypothesis is expanded upon in Chapter 3.

2.6 Conclusions

This study has improved our understanding of the complexity of phylogenetic relationships within the genus *Paphiopedilum* in general and section *Barbata* in particular. Further, it has uncovered phylogenetic signals pointing to widespread reticulate evolution within section *Barbata* and highlights the importance of using multiple samples and multiple loci for phylogenetic studies.

The morphologically unusual taxon *P. fairrieanum* from the Himalayas (Sikkim) was found to be misplaced in section *Paphiopedilum*. In this study, there is phylogenetic incongruence in the placement *P. fairrieanum* between the various loci, similar to the case of *P. canhii* in Górniak et al. (2014). Potentially *P. fairrieanum* should be placed into its own subgenus. More samples should be analysed to confirm this.

Section *Barbata* contains several complexes or groups of morphologically allied taxa that are of systematic interest. However, the correct resolution of these complexes is difficult due to conflicting phylogenetic signals and the lack of phylogenetic depth, possibly due to the youth of this lineage. Previous studies were unable to correctly treat these complexes due to low-taxon sampling or reliance on only a single locus that could not account for reticulation. For example, the ITS trees of Cox *et al*. (1997) and Chochai *et al*. (2012) showed support for the separation of *P. appletonianum* and *P. bullenianum* when the more likely interpretation is that there is reticulation. This study found various degrees of phylogenetic support for the *P. hookerae-sangii*, the *P. sukhakulii-venustum-wardii*, *P. mastersianum-papuanum-violascens*, *P. appletonianum-bullenianum* and *P. barbatum-callosum-lawrenceanum* alliances. Probably the circumscription of these taxa is confused through repeated reticulation.

To generate better understanding of the patterns of reticulation in this complex, it will be necessary to sample more taxa and use more markers in tree building. Potentially there will be value in sequencing genes that are involved in species isolation (e.g. flowering genes) and to apply population genetic markers from across the genome (e.g. RAD sequencing) for deeper insights. In Chapter 3, we complete a rigorous examination of the nature and occurrence of

Xdh paralogues, to better resolve patterns of reticulation in relation to phylogeography of the region.

Tables

Table 2.1 Details of samples and sequences used in this study

Sample	Area	Source	*ycf1*	*matK*	*trnF(GAA)- ndhJ*	*psaA-ycf3ex3*	*Xdh*	*CHS*	Verification (P/V)
Paphiopedilum									
section *Barbata*									
P. acmodontum A	F	MBG	This study	This study	NA	This study	This study	This study	P & V
P. acmodontum B	F	MBG	This study	This study	NA	This study	This study	This study	P & V
P. appletonianum A	B	RO	This study	This study	This study	This study	This study	This study	P
P. appletonianum B	B	K	This study	This study	This study	This study	This study	This study	V
P. appletonianum C	A	QM	This study	This study	This study	This study	This study	NA	P & V
P. appletonianum D	B	MBG	This study	This study	This study	This study	This study	NA	P & V
P. appletonianum E	B	BPS	This study	This study	This study	This study	This study	This study	P & V
P. appletonianum F	B	HG	This study	This study	This study	This study	This study	This study	P
P. appletonianum G	B	BPS	This study	This study	This study	This study	This study	This study	P
P. appletonianum H	A	QM	This study	This study	This study	This study	This study	This study	Pending
P. appletonianum I	A	QM	This study	This study	This study	This study	This study	This study	Pending
P. argus A	F	MBG	This study	This study	This study	This study	This study	NA	P & V
P. argus B	F	QM	This study	This study	This study	This study	This study	This study	Pending
P. barbatum A	C	K	This study	This study	This study	This study	This study	NA	P
P. barbatum B	C	K	This study	This study	This study	NA	This study	This study	P
P. barbatum C	C	RO	This study	This study	This study	This study	This study	This study	P
P. barbatum D	C	FRIM	This study	This study	This study	This study	This study	This study	P & V
P. barbatum E	C	FRIM	This study	This study	This study	This study	This study	This study	P & V
P. barbatum F	C	FRIM	This study	This study	This study	This study	This study	This study	P & V

(cont.)

Table 2.1 (cont.)

Sample	Area	Source	*ycf1*	*matK*	*trnF(GAA)- ndhJ*	*psaA-ycf3ex3*	*Xdh*	*CHS*	Verification (P/V)
section *Barbata (cont.)*									
P. barbatum G	C	FRIM	This study	This study	This study	This study	NA	This study	P & V
P. barbatum H	C	FRIM	This study	This study	This study	This study	This study	This study	P & V
P. barbatum I	C	FRIM	This study	NA	NA	This study	This study	This study	P & V
P. bullenianum A	C	K	This study	This study	This study	This study	This study	This study	P & V
P. bullenianum C	E	RO	This study	This study	This study	This study	This study	This study	P
P. bullenianum D	C	BPS	This study	This study	This study	This study	This study	This study	P
P. bullenianum E	C	QM	This study	This study	This study	This study	NA	This study	P & V
P. bullenianum F	D	BBG	This study	This study	This study	This study	This study	This study	Pending
P. bullenianum G†	G	K	This study	This study	This study	This study	This study	This study	-
P. bullenianum H	C	FRIM	This study	This study	This study	This study	This study	This study	P & V
P. bullenianum I	G	QM	This study	This study	NA	This study	This study	This study	Pending
P. bullenianum J	G	QM	This study	This study	This study	This study	This study	This study	Pending
P. bullenianum K	C	FRIM	This study	This study	This study	This study	This study	This study	V
P. bullenianum L	E	K	This study	This study	This study	This study	This study	This study	Pending
P. bullenianum M	C	K	This study	This study	This study	This study	This study	This study	P & V
P. bullenianum N†	E	K	This study	This study	This study	This study	This study	This study	-
P. bullenianum O	G	QM	This study	This study	This study	This study	NA	NA	Pending
P. callosum A	C	QM	This study	This study	This study	This study	This study	This study	Pending
P. callosum B	C	RO	This study	This study	This study	This study	This study	NA	P
P. callosum C	C	FRIM	This study	This study	This study	This study	This study	This study	P & V
P. callosum D	B	QM	This study	This study	This study	This study	This study	This study	P
P. callosum E	B	K	This study	NA	This study	This study	This study	This study	P

(cont.)

Table 2.1 (cont.)

Sample	Area	Source	*ycf1*	*matK*	*trnF(GAA)- ndhJ*	*psaA-ycf3ex3*	*Xdh*	*CHS*	Verification (P/V)
section *Barbata* (cont.)									
P. callosum F	B	K	This study	This study	This study	This study	This study	This study	V
P. callosum G	B	QM	This study	This study	This study	This study	This study	This study	Pending
P. ciliolare	F	QM	This study	This study	This study	This study	This study	This study	Pending
P. dayanum A	E	RO	This study	This study	This study	This study	This study	This study	P
P. dayanum B	E	KP	NA	NA	NA	NA	This study	NA	P
P. fowliei A	F	RO	This study	This study	This study	This study	This study	This study	P
P. fowliei B	F	K	This study	This study	This study	This study	This study	This study	V
P. hennisianum A	F	K	This study	This study	This study	This study	This study	This study	V
P. hennisianum B	F	RO	This study	This study	This study	This study	This study	This study	P
P. hookerae A†	E	QM	This study	This study	This study	This study	This study	This study	-
P. hookerae B	E	QM	This study	This study	This study	This study	This study	This study	P & V
P. hookerae C	E	KB	NA	NA	NA	NA	This study	NA	P
P. hookerae D	E	KB	NA	NA	NA	NA	This study	NA	P
P. inamorii	E	KP	NA	NA	NA	NA	This study	NA	P
P. javanicum A	D	RO	This study	This study	This study	This study	This study	This study	P
P. javanicum B	D	BBG	This study	This study	This study	This study	This study	This study	P
P. javanicum C	D	BBG	This study	This study	This study	This study	This study	This study	P
P. javanicum D	E	K	This study	This study	This study	This study	This study	NA	V
P. javanicum E	D	MBG	This study	This study	This study	This study	This study	This study	P & V
P. javanicum F	E	KB	NA	NA	NA	NA	This study	NA	P
P. lawrenceanum A	E	K	This study	This study	This study	This study	This study	NA	P
P. lawrenceanum B	E	HG	This study	This study	This study	This study	This study	This study	P

(cont.)

Table 2.1 (cont.)

Sample	Area	Source	*ycf1*	*matK*	*trnF(GAA)- ndhJ*	*psaA-ycf3ex3*	*Xdh*	*CHS*	Verification (P/V)
section *Barbata* (cont.)									
P. lawrenceanum C	E	RO	This study	This study	This study	This study	This study	This study	P
P. lawrenceanum D	E	K	This study	This study	This study	This study	This study	This study	P
P. lawrenceanum E	E	KP	NA	NA	NA	NA	This study	NA	P
P. lawrenceanum F	E	KB	NA	NA	NA	NA	This study	NA	P
P. mastersianum A	G	K	This study	This study	This study	This study	This study	This study	P & V
P. mastersianum B	G	K	This study	This study	This study	This study	This study	NA	Pending
P. mastersianum C	G	BBG	This study	This study	This study	This study	This study	This study	Pending
P. mastersianum D	G	MBG	NA	This study	This study	This study	NA	This study	P & V
P. mastersianum E	G	RO	This study	NA	This study	This study	This study	This study	P
P. papuanum A	H	QM	This study	This study	This study	This study	This study	This study	Pending
P. papuanum B	H	BBG	This study	This study	This study	This study	This study	This study	Pending
P. purpuratum	A	RO	This study	This study	This study	This study	This study	This study	P
P. sangii A	G	BBG	This study	This study	This study	This study	This study	This study	Pending
P. sangii B	G	BBG	This study	This study	This study	This study	This study	This study	Pending
P. sangii C	G	QM	This study	This study	This study	This study	This study	This study	Pending
P. schoseri	G	QM	This study	This study	This study	This study	This study	This study	P & V
P. sugiyamanum	E	KP	NA	NA	NA	NA	This study	NA	P
P. sukhakulii A	B	K	This study	This study	This study	This study	This study	This study	P
P. sukhakulii B	B	K	This study	This study	This study	This study	NA	This study	P
P. superbiens A	D	BBG	This study	This study	This study	NA	This study	This study	Pending
P. superbiens B	D	RO	This study	This study	This study	This study	This study	This study	P
P. superbiens C	D	K	This study	This study	This study	This study	This study	NA	V

(cont.)

Table 2.1 (cont.)

Sample	Area	Source	*ycf1*	*matK*	*trnF(GAA)- ndhJ*	*psaA-ycf3ex3*	*Xdh*	*CHS*	Verification (P/V)
section *Barbata* (cont.)									
P. superbiens D	D	RO	This study	This study	This study	This study	This study	This study	P
P. superbiens E	D	BBG	This study	NA	This study	This study	This study	This study	P
P. superbiens F	D	BBG	This study	This study	This study	This study	This study	This study	P
P. superbiens G	D	BBG	NA	NA	NA	NA	This study	NA	P
P. tonsum A	D	RO	NA	This study	This study	This study	This study	This study	P
P. tonsum B	D	QM	This study	This study	This study	This study	This study	This study	P
P. tonsum C	D	QM	This study	This study	This study	This study	This study	This study	P
P. tonsum D	D	K	This study	This study	This study	This study	This study	This study	P
P. tonsum E	D	BBG	This study	This study	This study	This study	This study	This study	P
P. tonsum F	D	BBG	This study	This study	This study	This study	This study	This study	P
P. urbanianum A	F	BPS	This study	This study	This study	This study	This study	This study	P
P. urbanianum B	F	QM	This study	This study	This study	This study	This study	This study	P & V
P. venustum A	A	RO	This study	This study	This study	This study	This study	This study	P
P. venustum B†	A	QM	This study	This study	This study	This study	This study	This study	-
P. venustum C	A	QM	This study	This study	This study	This study	This study	NA	P
P. venustum D	A	RO	This study	This study	This study	This study	This study	This study	P
P. violascens A	H	K	This study	This study	This study	This study	This study	This study	V
P. violascens B	H	QM	This study	This study	This study	This study	This study	NA	Pending
P. violascens C	H	QM	This study	This study	This study	This study	This study	This study	P & V
P. wardii A	A	K	This study	This study	This study	This study	This study	This study	P
P. wardii B	A	K	This study	This study	This study	This study	This study	This study	P
P. wardii C	A	K	NA	NA	NA	NA	NA	This study	P

(cont.)

Table 2.1 (cont.)

Sample	Area	Source	*ycf1*	*matK*	*trnF(GAA)- ndhJ*	*psaA-ycf3ex3*	*Xdh*	*CHS*	Verification (P/V)
section *Barbata* (cont.)									
P. × siamense†	B	K	This study	This study	This study	This study	This study	This study	-
subgenus *Brachypetalum*									
P. concolor	B	GenBank	JQ929520	JQ929367	JQ929469	JQ929418	This study	NA	V
P. niveum	B	K	JQ929544	JQ929391	JQ929493	JQ929442	This study	This study	V
section *Cochlopetalum*									
P. glaucophylum	D	BBG	This study	NA	NA	This study	This study	This study	P
P. liemianum	D	BPS	JQ929528	JQ929375	JQ929477	JQ929426	This study	This study	P
P. primulinum	D	K	This study	This study	This study	This study	This study	This study	P
P. victoria-mariae	D	K	NA	NA	NA	NA	This study	This study	Pending
P. victoria-regina	D	K	JQ929557	JQ929404	JQ929506	JQ929455	This study	This study	V
section *Coryopedilum*									
P. rothschildianum	E	K	JQ929550	JQ929397	JQ929499	JQ929448	This study	This study	V
P. stonei	E	K	JQ929553	JQ929400	JQ929502	JQ929451	This study	This study	V
P. supardii	E	BBG	This study	NA	This study	This study	This study	This study	P
subgenus *Megastaminodium*									
P. canhii	A	GenBank	NA	JQ660904	NA	NA	JQ660948	NA	

(cont.)

Table 2.1 (cont.)

Sample	Area	Source	*ycf1*	*matK*	*trnF(GAA)- ndhJ*	*psaA-ycf3ex3*	*Xdh*	*CHS*	Verification (P/V)
section *Paphiopedilum*									
P. coccineum	A	BPS	This study	This study	This study	This study	This study	This study	P
P. druryi		K	JQ929523	JQ929370	JQ929472	JQ929421	This study	NA	V
P. fairrieanum†	A	K	This study	This study	This study	This study	This study	This study	-
P. gratrixianum	A	K	JQ929529	JQ929376	JQ929478	JQ929427	NA	This study	V
P. spicerianum	A	K	JQ929552	JQ929399	JQ929501	JQ929450	This study	This study	V
section *Pardalopetalum*									
P. haynaldianum	C	K	JQ929532	JQ929379	JQ929481	JQ929430	This study	NA	V
P. lowii	C ?	K	JQ929540	JQ929387	JQ929489	JQ929438	This study	This study	V
P. parishii	B ?	GenBank	JQ929545	JQ929392	JQ929494	JQ929443	NA	NA	
subgenus *Parvisepalum*									
P. armeniacum	A	K	This study	This study	This study	This study	This study	This study	V
P. delenatii	A	K	JQ929521	JQ929368	JQ929470	JQ929419	This study	NA	V
P. hangianum	A	K	This study	This study	NA	This study	This study	This study	V
P. jackii†	A	QM	This study	This study	This study	This study	This study	This study	-
P. malipoense	A	K	JQ929541	JQ929388	JQ929490	JQ929439	This study	This study	V
P. micranthum	A	K	JQ929543	JQ929390	JQ929492	JQ929441	This study	NA	V
P. vietnamense A	A	BPS	This study	This study	This study	This study	This study	This study	Pending
P. vietnamense B	A	QM	This study	This study	This study	This study	This study	This study	Pending

(cont.)

Table 2.1 (cont.)

Mexipedium									
Mexipedium xerophyticum		GenBank	NA	NA	NA	NA	GU004510	NA	
Selenipedium									
Selenipedium aequinoctiale		GenBank	NA	NA	NA	NA	GU004507	NA	
Cypripedium									
Cypripedium passerinum		GenBank	NA	NA	NA	NA	GU004508	NA	
Cypripedium tibeticum		GenBank	NA	NA	NA	NA	GU004509	NA	

Abbreviations:

Area
A – North Indochina, South China & Himalayas
B – South Indochina
C – The Malaysian Peninsula
D – Sumatra & Java
E – Borneo
F – The Philippines
G – Wallacea
H – Papua New Guinea

Source
BBG – Bogor Botanic Gardens, Indonesia
BPS – British *Paphiopedilum* Society
FRIM – Forest Research Institute, Malaysia
HG – Herrenhausen Garten, Germany
K – Royal Botanic Gardens, Kew, UK
KB – Kinabalu Park, Malaysia
KP – Kipandi Park, Malaysia
MBG – Munich Botanic Gardens, Germany
QM – Queen Mary University of London, UK
RO – Ratcliffe Orchids, UK

Verification
V – Voucher
P – Photograph

Table 2.2 Details of primers used to amplify DNA regions. For details of the PCR mix (A-F) see Table 2.3 and for the PCR cycle programmes (A-F) see Table 2.4.

Primers	Sequence	Notes	PC	PCR	Reference
ycf1					
3720F	TACGTATGTAATGAACGAAT	PCR/Seq	A	A	Neubig et al. 2009
5500R	GCTGTTATTGGCATCAAACC	PCR/Seq	A	A	Neubig et al. 2009
560F	GATCTGGACCAATGCACATA	Seq			Neubig et al. 2009
850R	TTTGATTGGGATGATCCAAG	Seq			Neubig et al. 2009
matK					
390F	CGATCTATTCATTCAATATTT	PCR/Seq	B	B	Sun et al. 2001
1326R	TCTAGCACACGAAAGTCGAA	PCR/Seq	B	B	Sun et al. 2001
trnF(GAA)-ndhJ					
61L	CCTCGTGTCACCAGTTCAAA	PCR/Seq	C	C	Sun et al. 2001
62R	TGGATAGGCTGGCCCTTAC	PCR/Seq	C	C	Sun et al. 2001
psaAycf3ex ycf3ex3					
51L	GTTCCGGCGAACGAATAAT	PCR/Seq	C	C	Ebert and Peakall 2009
52R	GTCGGATCAAGCTGCTGAG	PCR/Seq	C	C	Ebert and Peakall 2009
Xdh					
X551F	GAAGAGCAGATTGAAGAW	PCR	D	D	Górniak et al. 2010
X1591R	AAYTGGAGCAACTCCACA	PCR	D	D	Górniak et al. 2010
Xp551F	GAAGAGCAGATTGAAGAAT	PCR/Seq	E	E	Górniak et al. 2010
Xp1590R	AAACTGGAGCRACTCCACCA	PCR/Seq	E	E	Górniak et al. 2010
Xdh1044F	ACTGATAGCCCTGCATGAGG	Seq			This study
Xdh1197R	TGTCACGCATGTACCTGAGC	Seq			This study
Xdh566F	TGCCTTTCTGGAAATTTATGC	Nested/S			This study
Xdh1588R	GAGCGACTCCACCATATACA	Nested/S			This study
CHS					
CHS20F	AGGAAGCGCCATTTTGTTTG	PCR/Seq	F	F	This study
CHS807R	ATGCTCTCAGCTTGTCCGGC	PCR/Seq	F	F	This study
CHS25F	GCGCCATTTTGTTTGGAATG	Seq			This study
Cloning					
T7	TAATACGACTCACTATAGGG				Commercial
SP6	CATTTAGGTGACACTATAG				Commercial

Table 2.3 PCR mixes (A-F, Phusion and Nested) used.

Mix A

PCR reagents	Vol (µL) per 25 µL of sample
dNTPs (10 mM)	0.50
5x GoTaq Flexi buffer	5.00
$MgCl_2$ (25 mM)	2.00
F primer (100 ng/µL)	0.34
R primer (100 ng/µL)	0.41
Go Taq (5 U)	0.10
BSA (0.4%)	0.00
DMSO	0.00
Template DNA*	1.00
dH_2O	15.65
Total	**25.00**

Mix B

PCR reagents	Vol (µL) per 40 µL of sample
dNTPs (10 mM)	0.80
5x GoTaq Flexi buffer	8.00
$MgCl_2$ (25 mM)	2.40
F primer (100 ng/µL)	2.00
R primer (100 ng/µL)	2.00
Go Taq (5 U)	0.80
BSA (4%)	4.00
DMSO	1.60
Template DNA*	1.00
dH_2O*	17.40
Total	**40.00**

Mix C

PCR reagents	Vol (µL) per 25 µL of sample
ReddyMix premix (Green cap, 2.5mM $MgCl_2$)	22.50
F primer (100 ng/µL)	0.50
R primer (100 ng/µL)	0.50
BSA (4%)	0.50
Template DNA*	1.00
Total	**25.00**

Table 2.3 (cont.)

Mix D

PCR reagents	Vol (µL) per 25 µL of sample
dNTPs (10 mM)	0.50
5x GoTaq Flexi buffer	5.00
$MgCl_2$ (25 mM)	1.00
F primer (100 ng/µL)	0.30
R primer (100 ng/µL)	0.30
Go Taq	0.25
BSA (4%)	0.90
DMSO	1.00
Template DNA*	1.00
dH_2O*	14.75
Total	**25.00**

Mix E

PCR reagents	Vol (µL) per 25 µL of sample
dNTPs (10 mM)	0.50
5x GoTaq Flexi buffer	5.00
$MgCl_2$ (25 mM)	2.00
F primer (100 ng/µL)	0.34
R primer (100 ng/µL)	0.30
Go Taq (5U)	0.20
DMSO	1.00
Template DNA*	1.00
dH_2O*	14.66
Total	**25.00**

Mix F

PCR reagents	Vol (µL) per 25 µL of sample
dNTPs	0.50
5xGoTag Flexibuffer	5.00
$MgCl_2$ (25 mM)	1.00
F primer (100 ng/µL)	0.30
R primer (100 ng/µL)	0.30
GoTaq (5U)	0.25
BSA (0.4%)	0.90
Template DNA*	1.00
dH_2O*	15.75
Total	**25.00**

* adjust DNA & dH_2O as necessary (use up to 4.0 µL of DNA for difficult samples)

Table 2.3 (cont.)

Phusion PCR Mix

PCR reagents	Vol (µL) per 25 µL of sample
dNTPs (10 mM)	0.50
5x Phusion HF buffer	5.00
F primer (100 ng/µL)	0.34
R primer (100 ng/µL)	0.30
Phusion polymerase (2U)	0.25
DMSO	1.00
Template DNA*	1.00
dH_2O*	16.61
Total	**25.00**

Nested PCR Mix

PCR reagents	Vol (µL) per 25 µL of sample
dNTPs (10 mM)	0.50
5x Phusion HF buffer	5.00
F primer (100 ng/µL)	0.35
R primer (100 ng/µL)	0.35
Phusion polymerase (2U)	0.25
DMSO	1.00
Template DNA*	1.00
dH_2O*	16.60
Total	**25.00**

Table 2.4 PCR cycling programmes (A-F, Phusion and Nested) used.

Cycle A

Temp. (°C)	Time (s)	
94	180	denaturation
94	30	touchdown: reducing 1°C each progressive cycle (8 cycles)
60-51	60	
72	180	
94	30	amplification (30 cycles: increase to 60 when necessary)
50	60	
72	180	
72	180	extension

Cycle B

Temp. (°C)	Time (s)	
94	120	denaturation
94	60	amplification (36 cycles)
53	60	
72	180	
72	320	extension

Cycle C

Temp. (°C)	Time (s)	
94	180	denaturation
94	60	amplification (28 cycles)
48	60	
72	60	
72	420	extension

Cycle D

Temp. (°C)	Time (s)	
94	120	denaturation
94	45	touchdown: reducing by 1°C each progressive cycle (6 cycles)
55-49	45	
72	90	
94	45	amplification (28 cycles)
49	45	
72	90	
72	300	extension

Table 2.4 (cont.)

Cycle E

Temp. (°C)	Time (s)	
94	120	denaturation
94	45	touchdown: reducing by 1°C each progressive cycle (6 cycles)
55-50	45	
72	90	
94	45	amplification (28 cycles: increase to 70 when necessary)
49	45	
72	90	
72	300	extension

Cycle F

Temp. (°C)	Time (s)	
94	120	denaturation
94	45	amplification (60 cycles)
58	45	
72	90	
72	300	extension

Phusion PCR cycle

Temp. (°C)	Time (s)	
98	120	denaturation
98	10	35 cycles
64	30	
72	30	
72	300	extension

Nested PCR cycle

Temp. (°C)	Time (s)	
98	120	denaturation
98	10	15 cycles
64	30	
72	30	
72	300	extension

Table 2.5 List of *Barbata* sequences included in the final concatenated dataset.

Combined plastid	***Xdh***	***CHS***
section *Barbata*		
P. acmodontum A	*P. acmodontum A*	*P. acmodontum A*
P. acmodontum B	*P. acmodontum B*	*P. acmodontum B*
P. appletonianum E	*P. appletonianum E2*	*P. appletonianum E*
P. appletonianum G	*P. appletonianum G*	*P. appletonianum G*
P. appletonianum F	*P. appletonianum F2*	*P. appletonianum F*
P. appletonianum I	*P. appletonianum I*	*P. appletonianum I*
P. appletonianum H	*P. appletonianum H*	*P. appletonianum H*
P. argus B	*P. argus B1*	*P. argus B*
P. barbatum H	*P. barbatum H*	*P. barbatum H*
P. barbatum F	*P. barbatum F*	*P. barbatum F*
P. barbatum C	*P. barbatum C*	*P. barbatum C*
P. bullenianum L	*P. bullenianum L1*	*P. bullenianum L*
P. bullenianum M	*P. bullenianum M*	*P. bullenianum M*
P. bullenianum G	*P. bullenianum G1*	*P. bullenianum G*
P. bullenianum C	*P. bullenianum C*	*P. bullenianum C*
P. bullenianum N	*P. bullenianum N3*	*P. bullenianum N*
P. bullenianum K	*P. bullenianum K*	*P. bullenianum K*
P. bullenianum J	*P. bullenianum J*	*P. bullenianum J*
P. violascens C	*P. violascens C*	*P. violascens C*
P. callosum F	*P. callosum F*	*P. callosum F*
P. callosum C	*P. callosum C*	*P. callosum C*
P. callosum D	*P. callosum D*	*P. callosum D*
P. ciliolare	*P. ciliolare*	*P. ciliolare*
P. superbiens D	*P. superbiens D4*	*P. superbiens D*
P. dayanum A	*P. dayanum A1*	*P. dayanum A*
P. fowliei A	*P. fowliei A*	*P. fowliei A*
P. fowliei B	*P. fowliei B1*	*P. fowliei B*
P. hennisianum A	*P. hennisianum A*	*P. hennisianum A*
P. hennisianum B	*P. hennisianum B*	*P. hennisianum B*
P. hookerae B	*P. hookerae B*	*P. hookerae B*
P. javanicum C	*P. javanicum C*	*P. javanicum C*
P. javanicum E	*P. javanicum E*	*P. javanicum E*
P. javanicum A	*P. javanicum A*	*P. javanicum A*
P. lawrenceanum A	*P. lawrenceanum A*	*P. lawrenceanum A*
P. lawrenceanum C	*P. lawrenceanum C*	*P. lawrenceanum C*
P. mastersianum A	*P. mastersianum A3*	*P. mastersianum A*
P. mastersianum C	*P. mastersianum C*	*P. mastersianum C*
P. mastersianum E	*P. mastersianum E1*	*P. mastersianum E*
P. papuanum B	*P. papuanum B*	*P. papuanum B*
P. papuanum A	*P. papuanum A1*	*P. papuanum A*
P. purpuratum	*P. purpuratum 3*	*P. purpuratum*
P. sangii C	*P. sangii C2*	*P. sangii C*
P. schoseri	*P. schoseri*	*P. schoseri*
P. superbiens A	*P. superbiens A*	*P. superbiens A*
P. superbiens B	*P. superbiens B*	*P. superbiens B*
P. tonsum D	*P. tonsum D*	*P. tonsum D*

(cont.)

Table 2.5 (cont.)

Combined plastid	*Xdh*	*CHS*
section *Barbata* (cont.)		
P. tonsum E	*P. tonsum E*	*P. tonsum E*
P. tonsum F	*P. tonsum F*	*P. tonsum F*
P. urbanianum A	*P. urbanianum A*	*P. urbanianum A*
P. urbanianum B	*P. urbanianum B1*	*P. urbanianum B*
P. venustum B	*P. venustum B*	*P. venustum B*
P. venustum D	*P. venustum D*	*P. venustum D*
P. violascens A	*P. violascens A*	*P. violascens A*
P. wardii A	*P. wardii A1*	*P. wardii A*
P. wardii B	*P. wardii B1*	*P. wardii B*
subgenus Brachypetalum		
P. niveum	*P. niveum*	*P. niveum*
section *Cochlopetalum*		
P. glaucophyllum	*P. glaucophyllum*	*P. glaucophyllum*
P. liemianum	*P. liemianum*	*P. liemianum*
P. primulinum	*P. primulinum*	*P. primulinum*
P. victoria-mariae	*P. victoria-mariae*	*P. victoria-mariae*
P. victoria-regina	*P. victoria-regina*	*P. victoria-regina*
section *Coryopedilum*		
P. rothschildianum	*P. rothschildianum*	*P. rothschildianum*
P. stonei	*P. stonei*	*P. stonei*
P. supardii	*P. supardii*	*P. supardii*
section *Paphiopedilum*		
P. coccineum	*P. coccineum*	*P. coccineum*
P. spicerianum	*P. spicerianum*	*P. spicerianum*
section *Pardalopetalum*		
P. lowii	*P. lowii*	*P. lowii*
subgenus *Parvisepalum*		
P. armeniacum	*P. armeniacum*	*P. armeniacum*
P. jackii	*P. jackii*	*P. jackii*
P. malipoense	*P. malipoense*	*P. malipoense*
P. vietnamense	*P. vietnamense*	*P. vietnamense*
Phragmipedium		
Phragmipedium besseae	*Phragmipedium besseae*	*Phragmipedium besseae*

Table 2.6 Summary of alignment lengths and number of characters in the plastid, nuclear *Xdh* and *CHS* datasets.

	ycf1	***matK***	***trnF(GAA)-ndhJ***	***psaA-ycf3ex3***	***Xdh***	***CHS***
Alignment length (bp)	1855	809	629	802	1005	736
No. of variable characters	167	100	70	67	264	141
No. of parsimonious characters	76	51	41	33	160	67

Figures

A

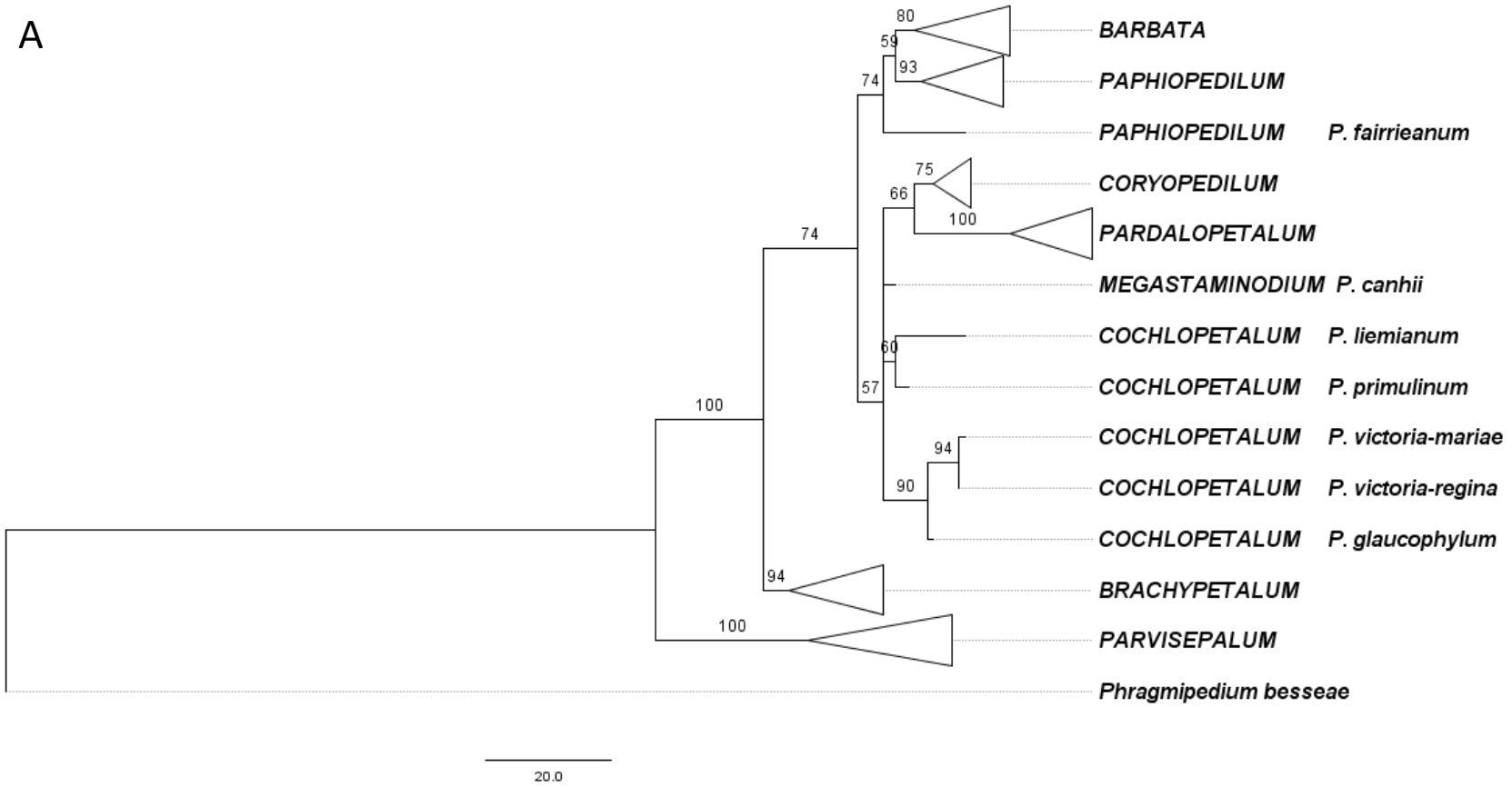

Figure 2.1 Trees showing the subgenera/section relationships in *Paphiopedilum* produced by **(A)** MP and **(B)** Bayesian analysis of the combined plastid coding *ycf1*, *matK*, and non-coding *psa-ycf3ex3* and *trnF(GAA)-ndhJ* regions. Bootstrap percentages (BP) for MP above 50 and posterior probabilities (PP) greater than 0.5 for Bayesian analysis are indicated above tree nodes.

(cont.)

B

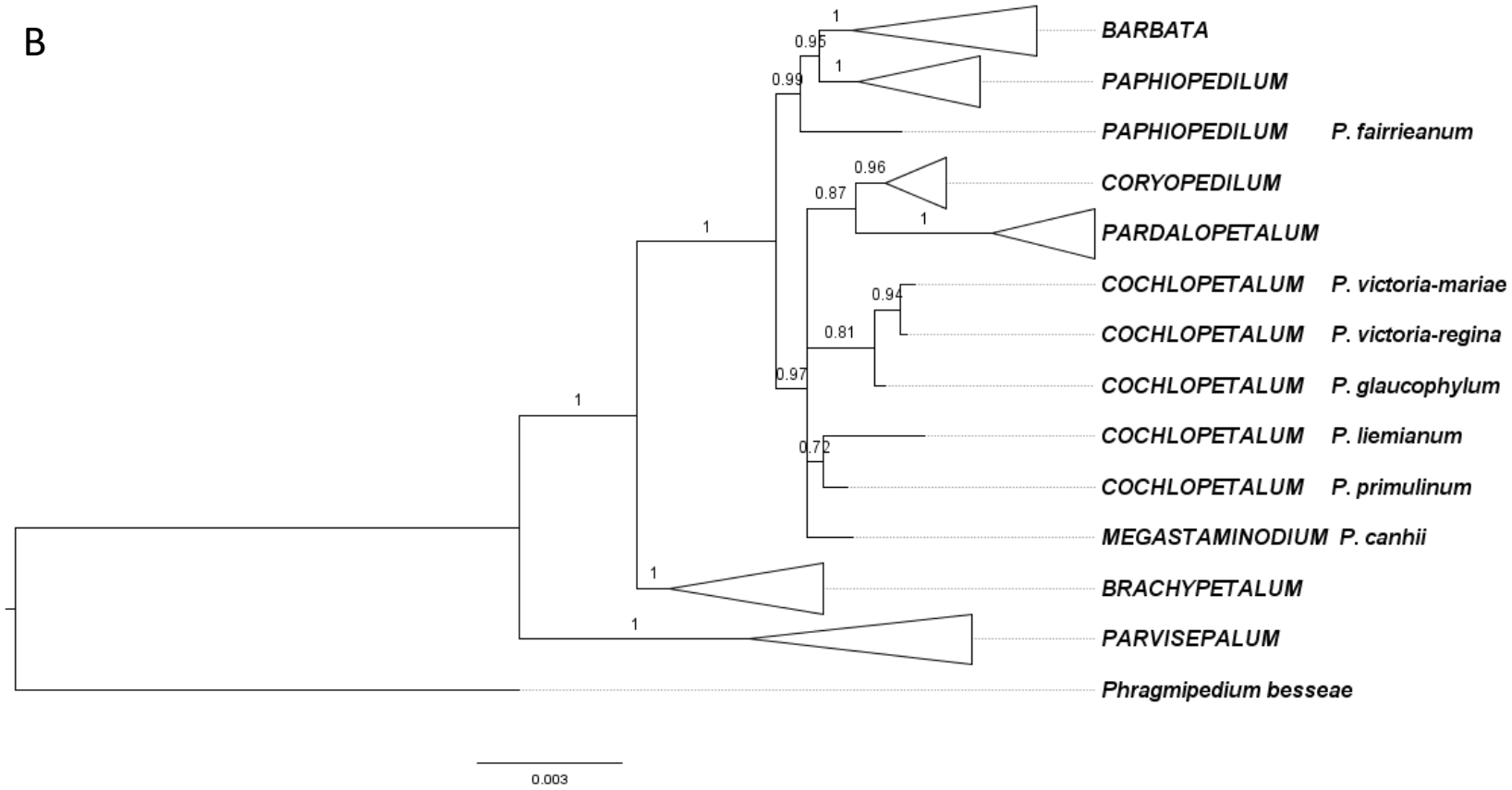

Figure 2.1 (cont.)

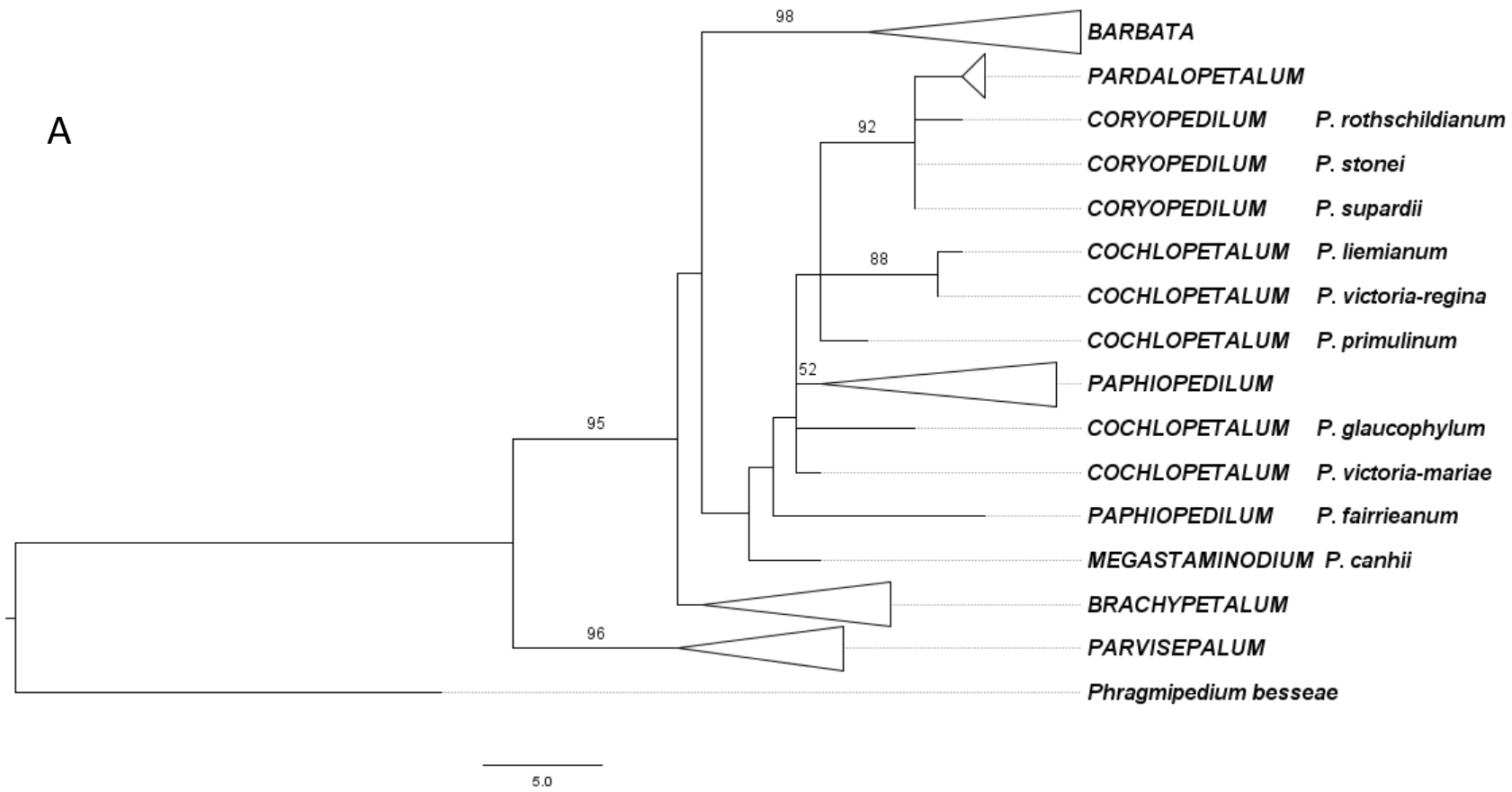

Figure 2.2 Trees showing the subgenera/section relationships in *Paphiopedilum* produced by (A) MP and (B) Bayesian analysis of the nuclear *Xdh* region. Bootstrap percentages (BP) for MP above 50 and posterior probabilities (PP) greater than 0.5 for Bayesian analysis are indicated above tree nodes.

(cont.)

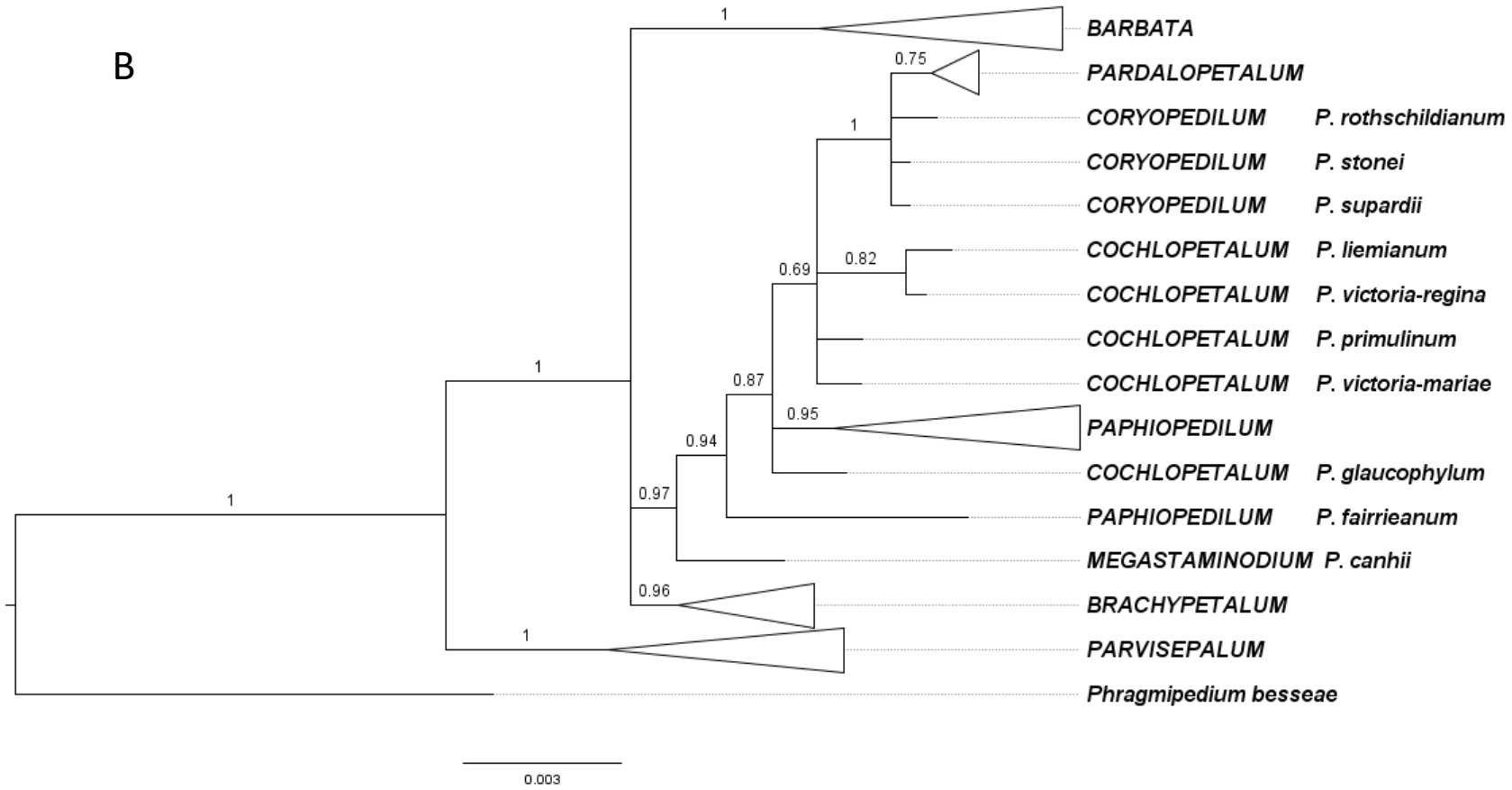

Figure 2.2 (cont.)

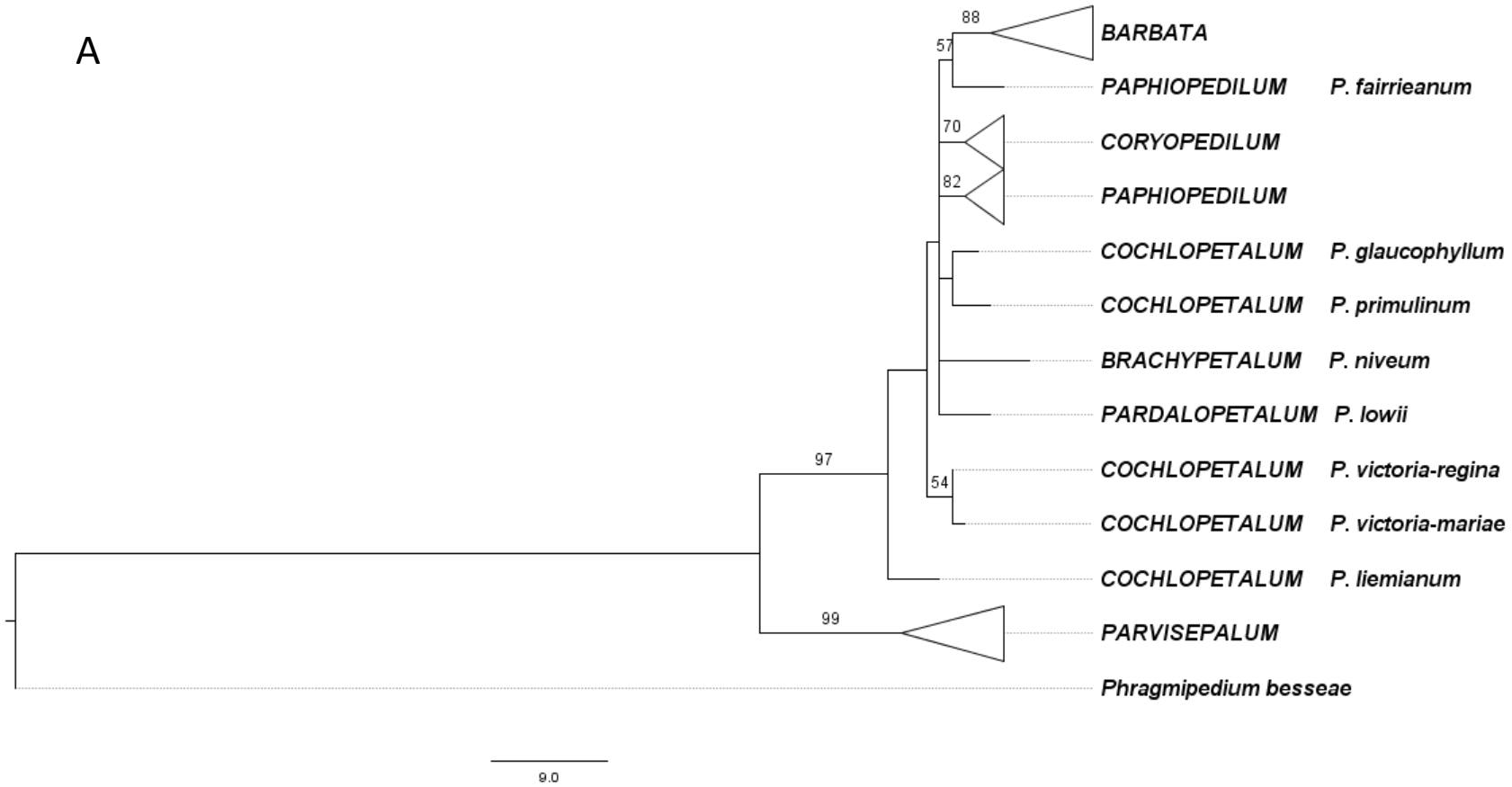

Figure 2.3 Trees showing the subgenera/section relationships in *Paphiopedilum* produced by (A) MP and (B) Bayesian analysis of the nuclear CHS region. Bootstrap percentages (BP) for MP above 50 and posterior probabilities (PP) greater than 0.5 for Bayesian analysis are indicated above tree nodes.

(cont.)

B

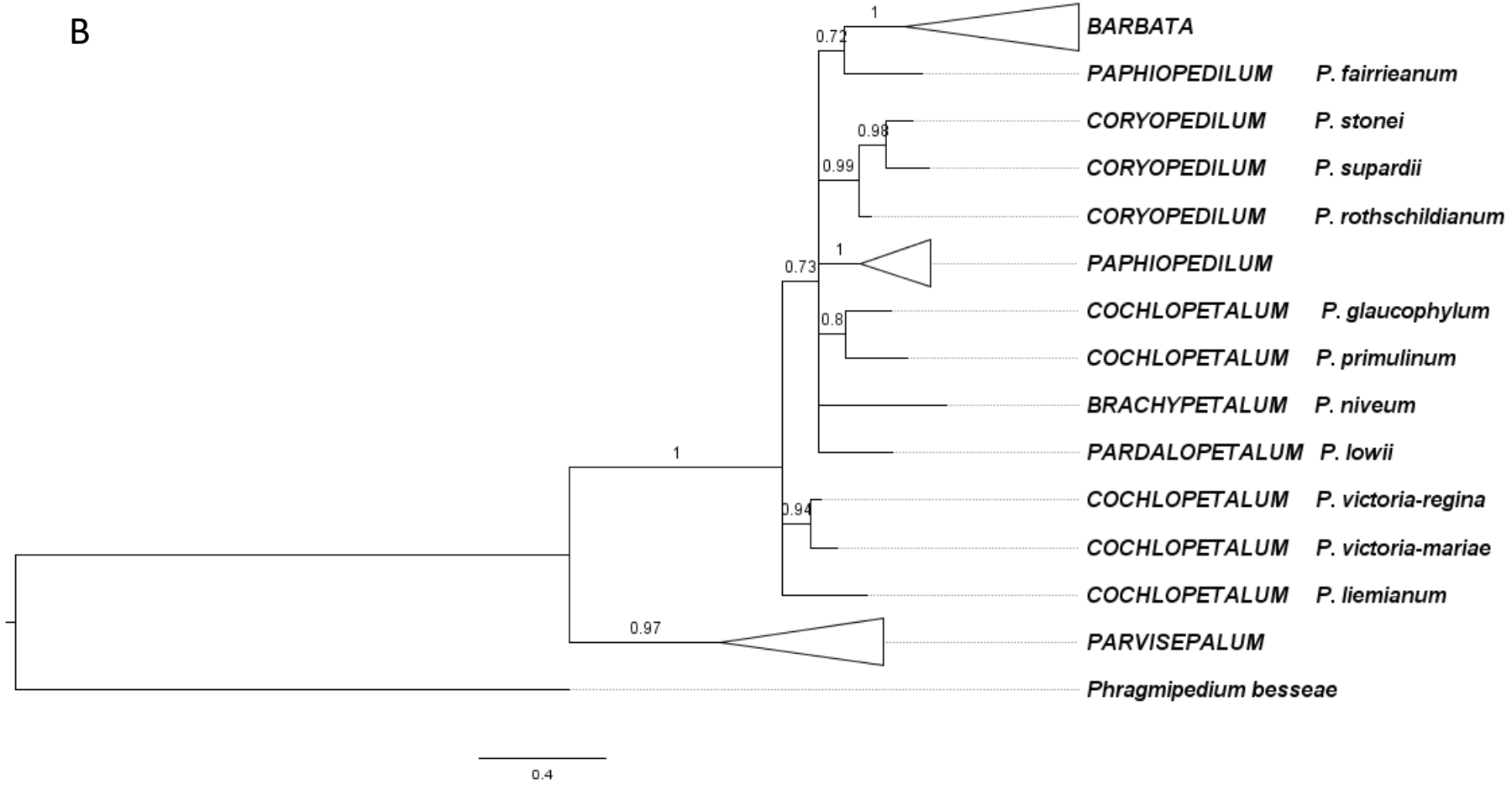

Figure 2.3 (cont.)

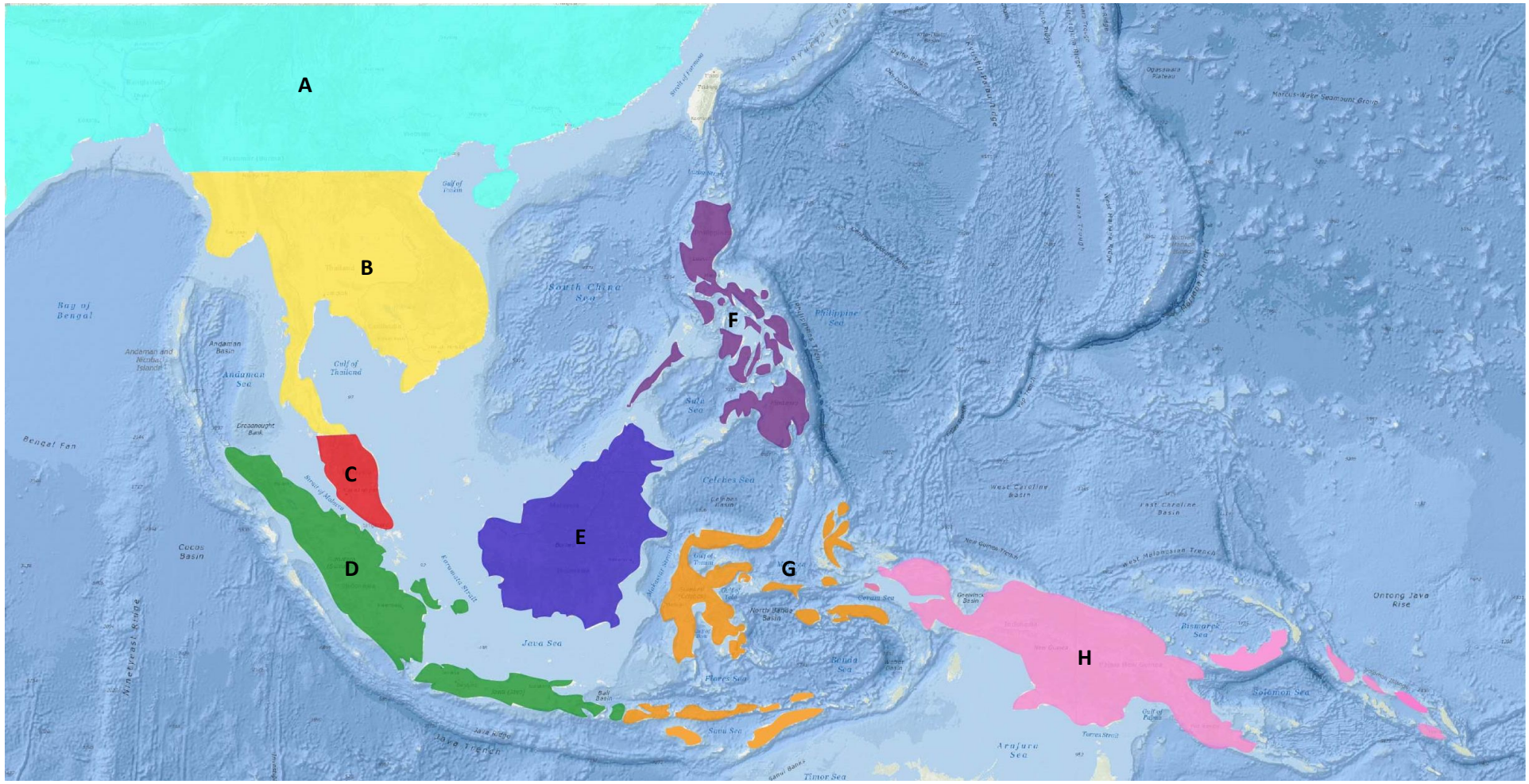

Figure 2.4 Map of Southeast Asia showing the geographical regions discussed in this study. Coloured areas are (**A**) North Indochina, South China & Himalayas, (**B**) South Indochina, (**C**) The Malaysian Peninsula, (**D**) Sumatra & Java, (**E**) Borneo, (**F**) The Philippines, (**G**) Wallacea, and (**H**) Papua New Guinea and correspond to the distributions of taxa samples in Fig. 2.5-2.8 & 3.1.

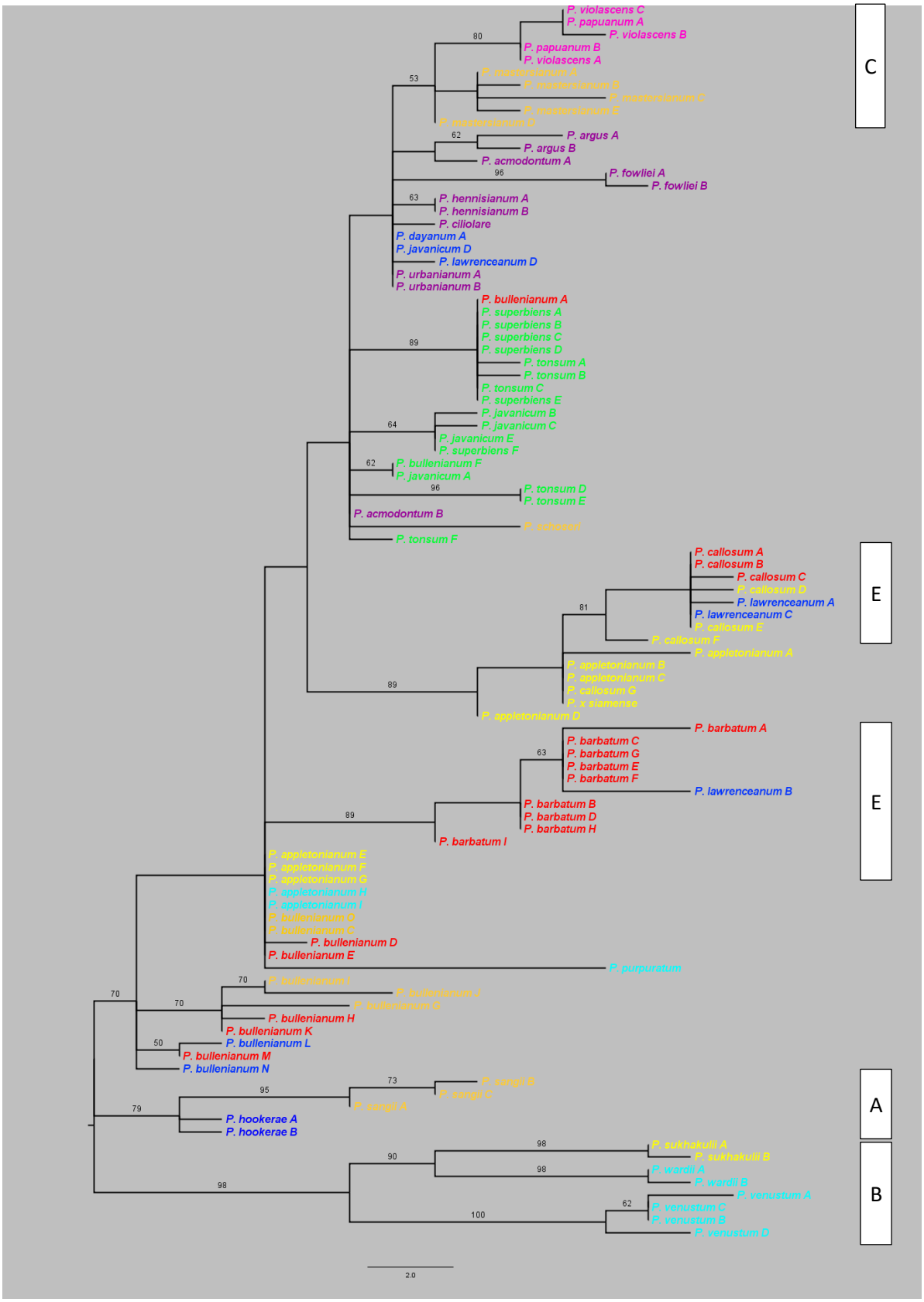

Figure 2.5A A tree showing the taxon relationships within section *Barbata* produced by MP analysis of the combined plastid coding *ycf1*, *matK*, and non-coding *psa-ycf3ex3* and *trnF(GAA)-ndhJ* regions. Numbers above ranches indicate bootstrap (BP) percentages above 50. The colours of the tip names indicate their origin and correspond to the geographical ranges displayed in Fig. 2.4. Bars indicate the major species groupings discussed in this study (**A**=*hookerae-sangii*; **B**=*sukhakulii-venustum-wardii*; **C**=*mastersianum-papuanum-violascens*; **D**=*appletonianum-bullenianum*; **E**=*barbatum-callosum-lawrenceanum*).

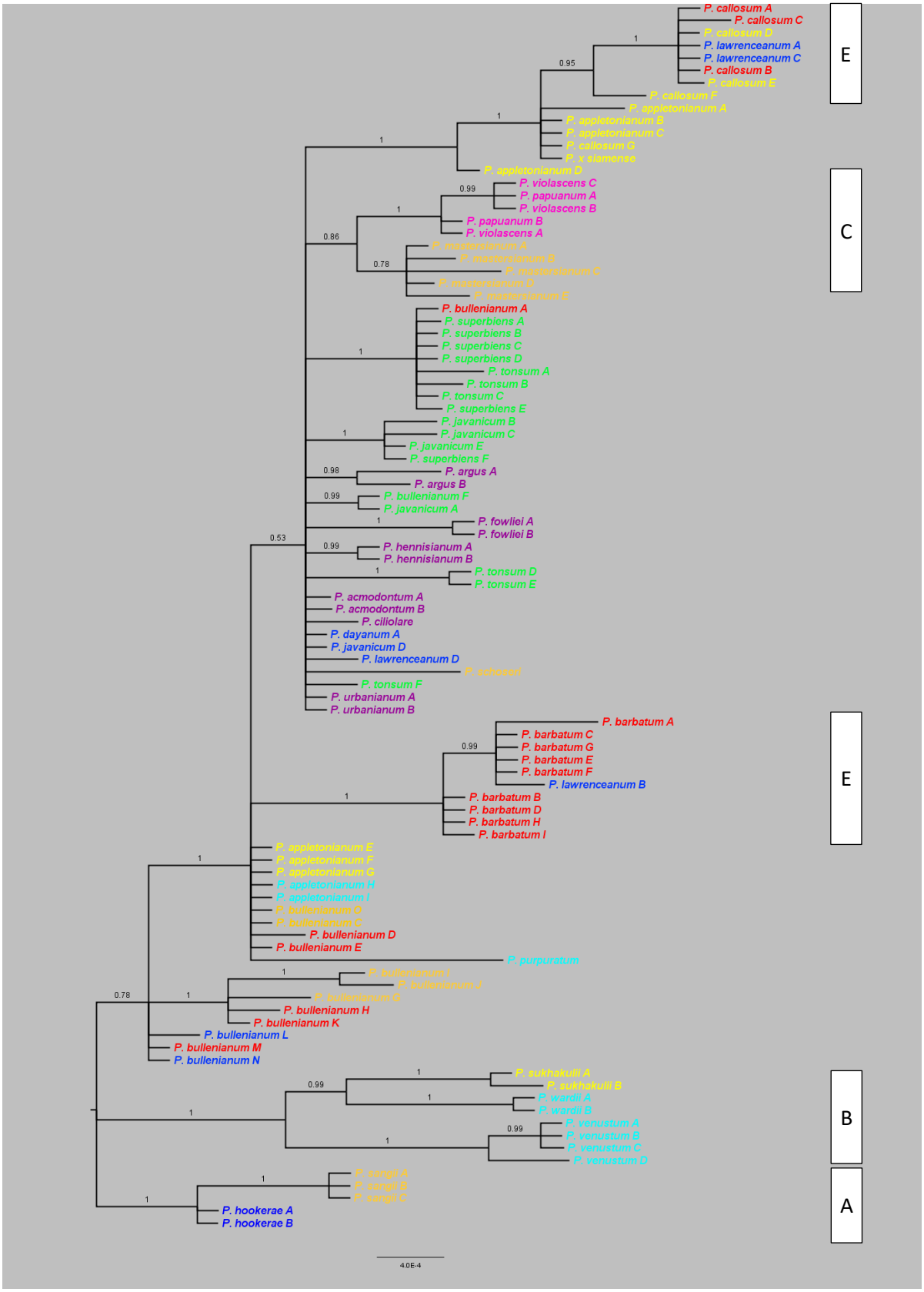

Figure 2.5B A tree showing the taxon relationships within section *Barbata* produced by Bayesian analysis of the combined plastid coding *ycf1*, *matK*, and non-coding *psa-ycf3ex3* and *trnF(GAA)-ndhJ* regions. Numbers above branches indicate posterior probability (PP) scores above 0.5. The colours of the tips indicate their origin correspond to the geographical ranges displayed in Fig. 2.4. Bars indicate the major species groupings discussed in this study (**A**=*hookerae-sangii*; **B**=*sukhakulii-venustum-wardii*; **C**=*mastersianum-papuanum-violascens*; **D**=*appletonianum-bullenianum*; **E**=*barbatum-callosum-lawrenceanum*).

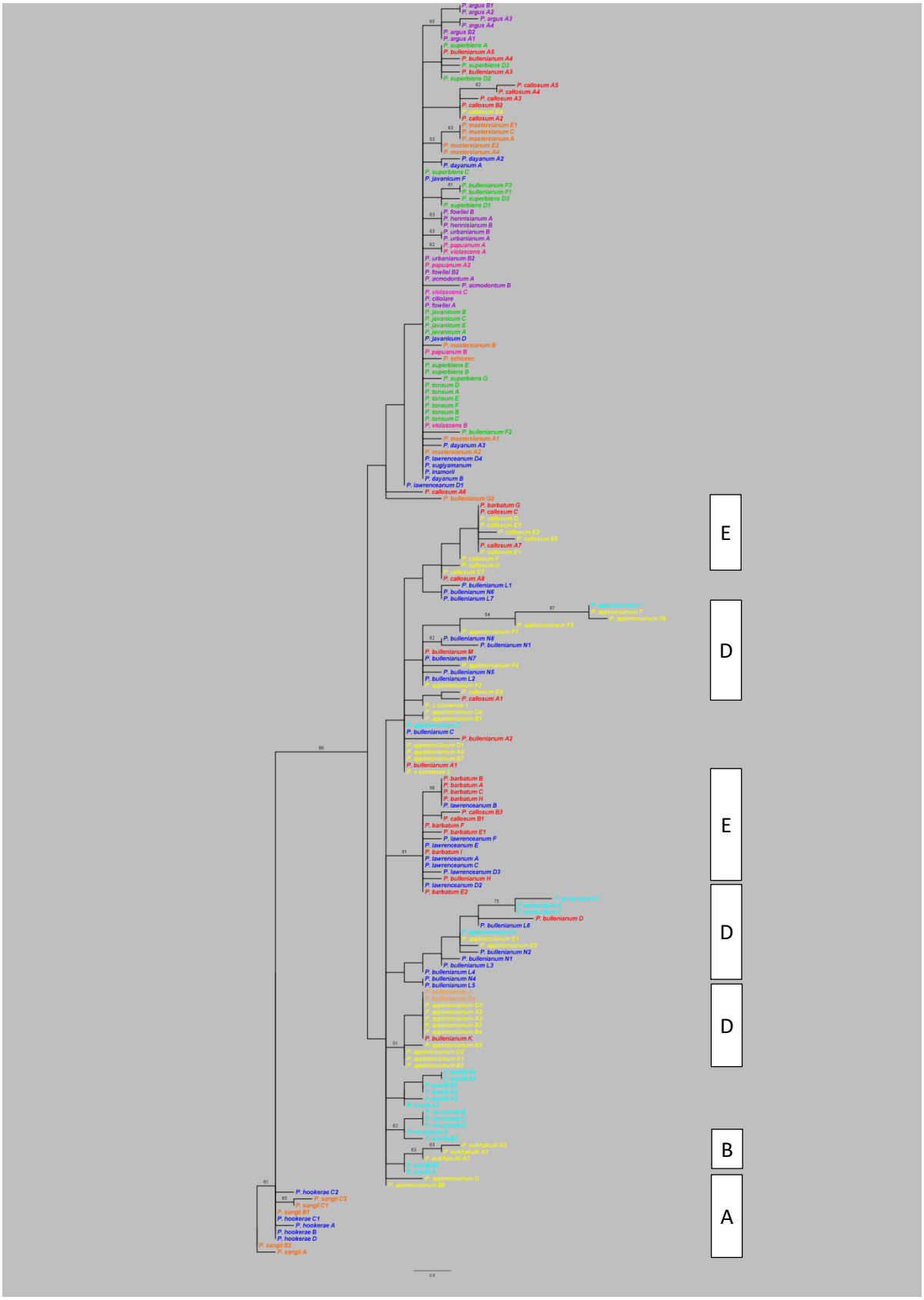

Figure 2.6A A tree showing the taxon relationships within section *Barbata* produced by MP analysis of the cloned nuclear *Xdh* region. Numbers above branches indicate bootstrap (BP) percentages above 50. The colours of the tips indicate their origin and correspond to the geographical ranges displayed in Fig. 2.4. Bars indicate clades containing the major species groupings discussed in this study (**A**=*hookerae-sangii*; **B**=*sukhakulii-venustum-wardii*; **C**=*mastersianum-papuanum-violascens*; **D**=*appletonianum-bullenianum*; **E**=*barbatum-callosum-lawrenceanum*).

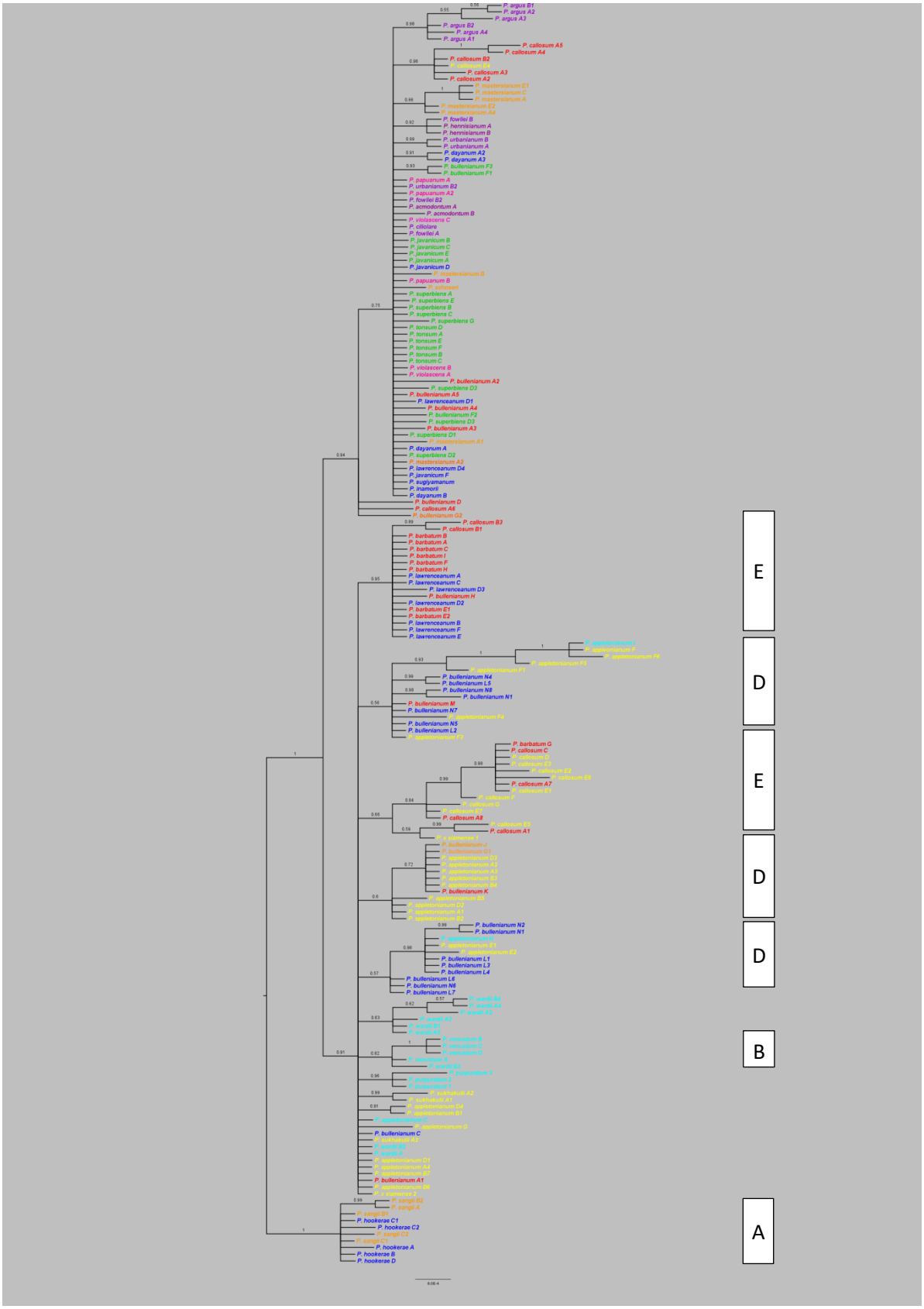

Figure 2.6B A tree showing the taxon relationships within section *Barbata* produced by Bayesian analysis of the cloned nuclear *Xdh* region. Numbers above branches indicate posterior probability (PP) scores above 0.5. The colours of the tips indicate its origin and correspond to the geographical ranges displayed in Fig. 2.4. Bars indicate clades containing the species groupings discussed in this study (**A**=*hookerae-sangii*; **B**=*sukhakulii-venustum-wardii*; **C**=*mastersianum-papuanum-violascens*; **D**=*appletonianum-bullenianum*; **E**=*barbatum-callosum-lawrenceanum*).

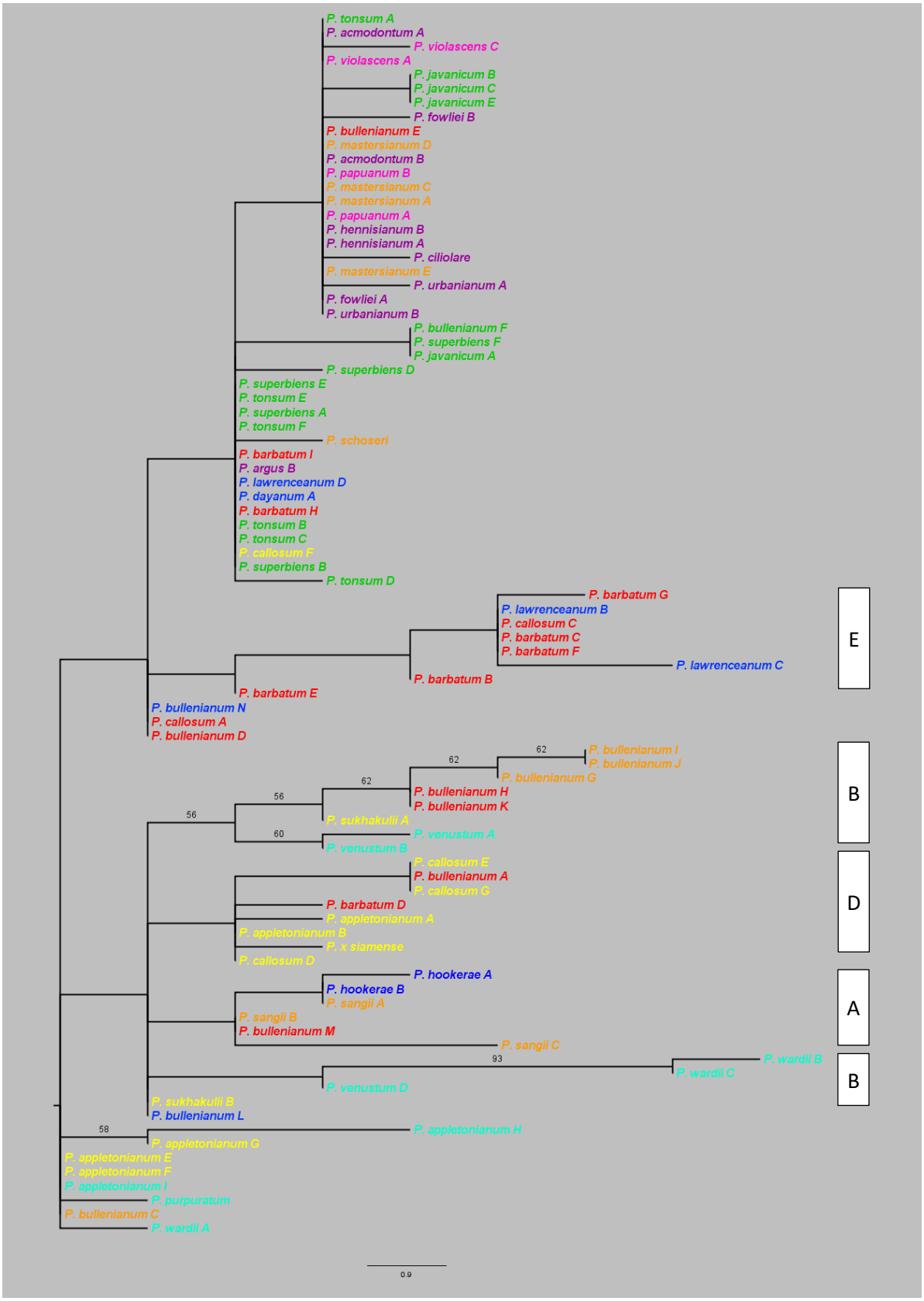

Figure 2.7A A tree showing the taxon relationships within section *Barbata* produced by MP analysis of the nuclear *CHS* region. Numbers above branches indicate bootstrap (BP) percentages above 50. The colours of the tips indicate their origin and correspond to the geographical ranges displayed in Fig. 2.4. Bars indicate clades containing the species groupings discussed in this study (**A**= *hookerae-sangii*; **B**=sukhakulii-*venustum-wardii*; **C**=*mastersianum-papuanum-violascens*; **D**=*appletonianum-bullenianum*; **E**=*barbatum-callosum-lawrenceanum*).

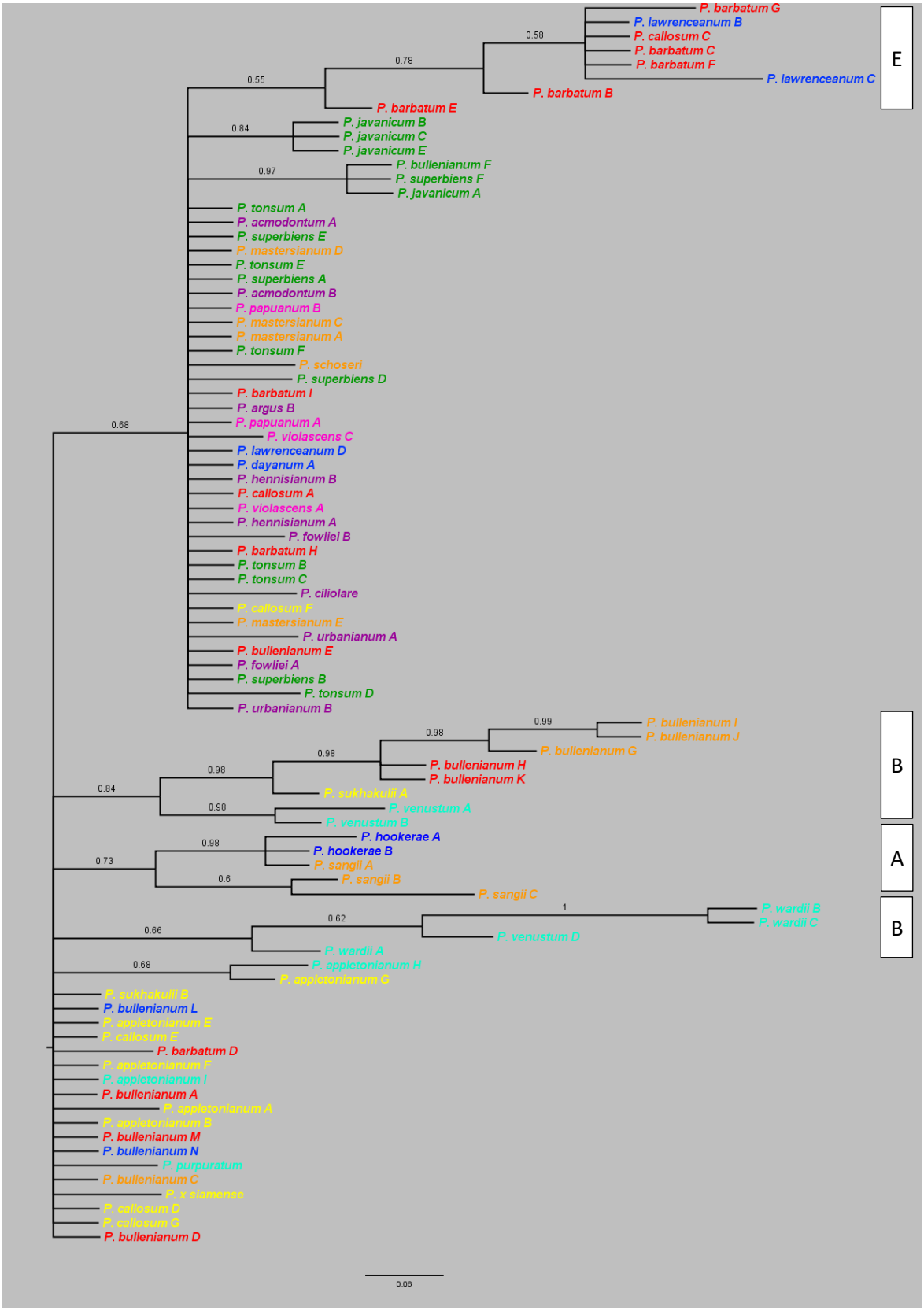

Figure 2.7B A tree showing the taxon relationships within section *Barbata* produced by Bayesian analysis of the nuclear *CHS* region. Numbers above branches indicate posterior probability (PP) scores above 0.5. The colours of the tips indicate their origin and correspond to the geographical ranges displayed in Fig. 2.4. Bars indicate clades containing the species groupings discussed in this study (**A**=*hookerae-sangii*; **B**=*sukhakulii-venustum-wardii*; **C**=*mastersianum-papuanum-violascens*; **D**=*appletonianum-bullenianum*; **E**=*barbatum-callosum-lawrenceanum*).

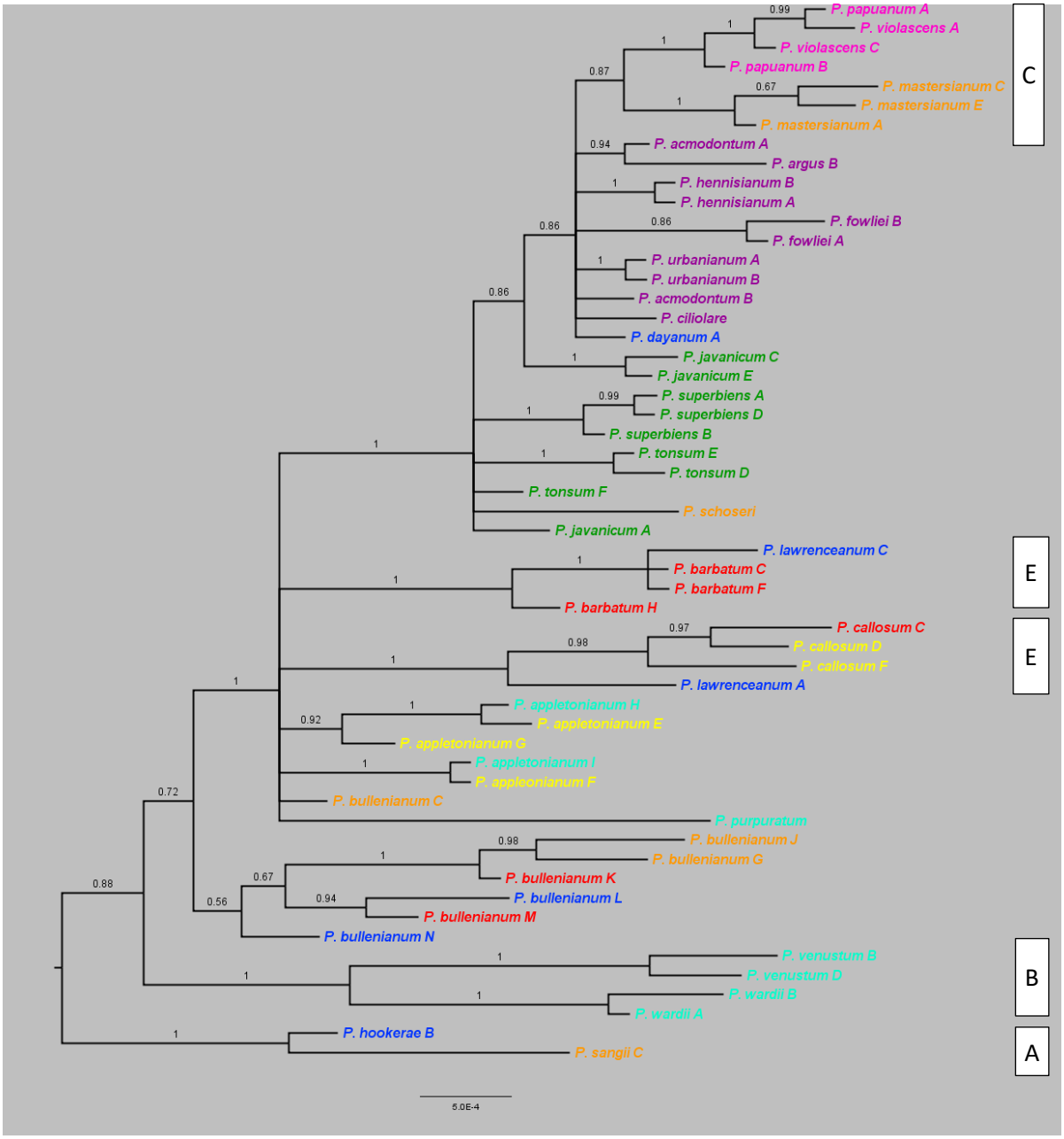

Figure 2.8 A tree showing the taxon relationships within section *Barbata* produced by Bayesian analysis of the concatenated plastid, cloned *Xdh* and *CHS* regions. Numbers above branches indicate posterior probability (PP) scores above 0.5. The colours of the tips indicate their origin and correspond to the geographical ranges displayed in Fig. 2.4. Bars indicate clades containing the species groupings discussed in this study (**A**=*hookerae-sangii*; **B**=*sukhakulii-venustum-wardii*; **C**=*mastersianum-papuanum-violascens*; **D**=*appletonianum-bullenianum*; **E**=*barbatum-callosum-lawrenceanum*).

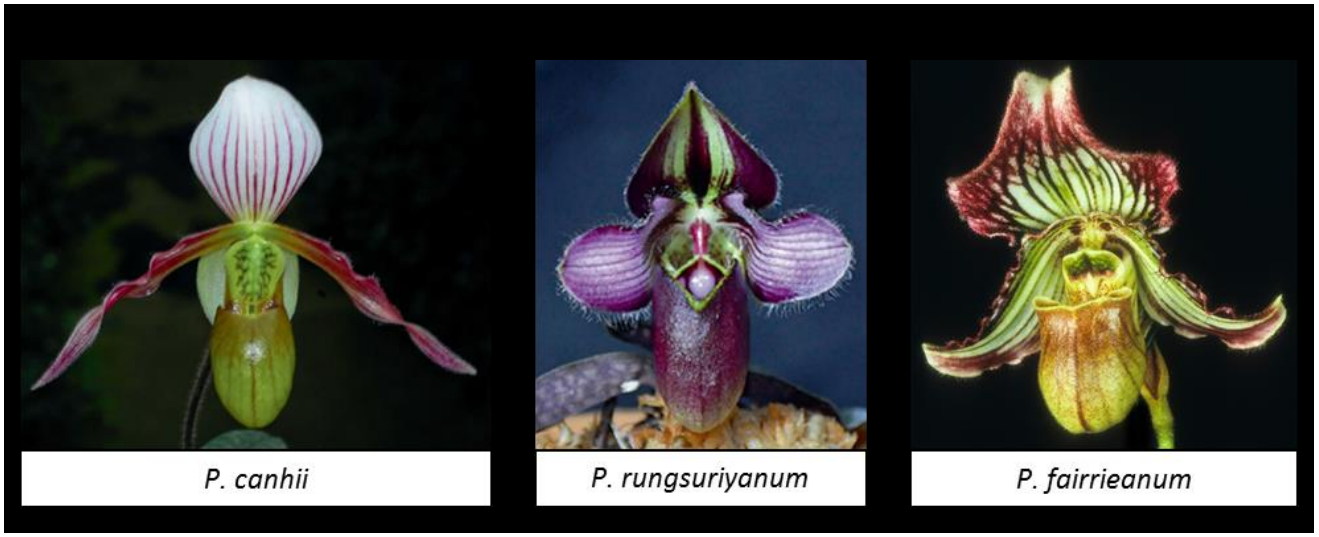

Figure 2.9 *Paphiopedilum canhii*, *P. rungsuriyanum* and *P. fairrieanum*. These three taxa have unusual morphologies that fall outside of the established subgenera/sections of Cribb (1998). Photo credits: *P. canhii* (Gruss O), *P. rungsuriyanum* (Gruss O) and *P. fairrieanum* (Levy J).

CHAPTER 3
MOLECULAR DATING AND BIOGEOGRAPHY

3.1 Abstract

Phylogenetic analyses of the cloned low-copy nuclear gene *Xdh* recovered three major clades: A, B1 and B2. Clade A is strongly supported and composed of taxa with a chromosomal complement of 2*n*=28, namely *P. hookerae* and *P. sangii*, and is sister to all other members in section *Barbata* (2n=32-42). Molecular dating estimates that this clade diverged around 6.3 Mya (95% HPD range of 4.0 to 8.8 Mya). The remaining two sister clades B1 and B2 show strong geographic signatures and are respectively composed of taxa from continental SE Asia (N & S Indochina and the Malaysian Peninsula) and island SE Asia (Borneo, Java, Sumatra, Wallacea, Philippines and Papua New Guinea) with a stem age of 5.1 Mya (95% HPD range of 3.4 to 7.2 Mya). Analysis of the cloned *Xdh* sequences has provided additional evidence of historical hybridisation in *Paphiopedilum* section *Barbata*. A Lagrange biogeographical reconstruction on a concatenated plastid, *Xdh* and *CHS* dataset suggests that *Paphiopedilum* section *Barbata* dispersed from continental SE Asia across Sunda and into Sahul. The signals further suggest repeated biotic exchanges across Sunda, especially between Borneo and continental SE Asia, and support the hypothesis of refugia on Borneo, the Malaysian Peninsula and South Indochina. Collectively, these analyses suggest that diversification in section *Barbata* is primarily driven by hybridisation, vicariance and dispersal facilitated by the glacial-interglacial cycles of sea-level fluctuations in SE Asia, and possibly chromosomal changes.

3.2 Introduction

SE Asia, also known as Malesia, is a fascinating biogeographical region composed of continental and island elements. Its greater archipelago is comprised of islands that are oceanic (e.g. most of the Philippines and Wallacea), continental crust or terranes (e.g. Palawan), or of composite (e.g. Sulawesi and Papua New Guinea) origin. The ancient movement of continental shelves and their associated terranes, if they remain above sea level, is important for understanding the biogeography of the region as they can function as rafts that transport biota from one continental shelf to another. SE Asia is the site of several continental plate collisions (reviewed in Lohman et al. 2011), the most significant of which is

probably the Sunda-Sahul collision (23-25 Mya) which brought about the emergence of landmasses and introduced Australian biotic elements into the region. This collision is also believed to be indirectly responsible for facilitating the emergence of rainforest biota by altering the prevailing oceanic currents and causing the annual monsoon (Morley 2012).

Glacial and inter-glacial cycles have repeatedly caused global fluctuations in sea-levels (Fig. 3.1), which are thought to have significantly influenced the biogeography of the area (e.g. Guo et al. 2015). During the past 2.7 Myr, sea-levels would have fallen by up to 120 m below present day levels for substantial periods of time (Miller et al. 2011). Each glacial maximum or ice-age would have been accompanied by falling of sea-levels, resulting in an increase in size of landmasses. This in turn would have facilitated colonisation, not only by exposing land bridges, but also by narrowing distances between islands and providing larger targets for dispersal. This effect would have been especially pronounced in SE Asia which lies partially on the shallow Sunda shelf. At its extreme, the falling sea-levels would have resulted in the connection of South Indochina and the Malaysian Peninsula with the present day islands of Sumatra, Java and Borneo to form a large peninsula known as Sundaland. This phenomenon is thought to account for the great biotic similarity between the Sunda landmasses, which is absent in the non-Sunda islands of Wallacea, the Philippines and Papua New Guinea which are on the Sahul shelf. The difference between land masses on the Sunda and Sahul shelves is especially marked in the fauna of the region which display a sharp transition across the hypothetical Wallace Line. Plants however may show a more gradual transitioning, probably due to their superior dispersal ability (Richardson et al. 2012).

Paleoclimatic conditions are another important aspect in understanding the biogeographical history of the region as biota can only make use of land connections if suitable habitat is present. The annual monsoon cycle began around 25 Mya and is believed to be a major contributing factor that led to overall wetter climates that supported the emergence of rainforest biota in SE Asia (Morley 2000). Climatic conditions would have varied over such timeframes. During the Pleistocene interglacials, the climate is thought to have been generally cooler and drier than the present day (Morley 2012). Under these conditions, it is thought that rainforests, though covering an overall larger area than at present, would have occurred as blocks centered on the mountains of northern Borneo and western Sumatra (Morley 2000; Cannon et al. 2009; Raes et al. 2014). However, the extent and coverage of Sundaland forests and the degree of connectivity between the blocks during glacial maxima is not known (Cannon et al. 2009).

From the above, it is clear that the historical biogeography of SE Asia is complex with different forces, including plate tectonics and changing sea levels, having had prominent roles at different times. Consequently, in order to reconstruct the biogeography of an area, it is essential that an estimate of the likely age of a lineage under study is established. In this chapter, I examine the biogeography of *Paphiopedilum* section *Barbata*, which we know is a herbaceous rainforest specialist, the present range of which encompasses the whole of SE Asia.

To address the overarching question "What biogeographical forces are driving the diversification of *Barbata* in southeast Asia?" specifically, I ask: (1) when did section *Barbata* diversify? (2) where did it originate? (3) what was its colonisation route? (4) what do the phylogenetic signals tell us about the biogeographical history of the region? and (5) do the new data support or refute any current theories relating to regional biogeography?

3.3 Materials & Methods

3.3.1 Molecular dating

Divergence times of species in section *Barbata* were estimated in two stages. In the first stage, a fossil-calibrated plastid *matK-rbcL* dataset for Orchidaceae was used to derive the crown age of *Paphiopedilum*. This result was then used as a secondary calibration point on the cloned low-copy nuclear *Xdh* dataset to estimate the divergence ages of species in section *Barbata*.

The combined partitioned *matK-rbcL* dataset consisted of sequences from Guo et al. (2012) with additional *Paphiopedilum* sequences (Table 3.3) derived from the methods detailed in Chapter 4. Calibration points include the fossils of subtribe Goodyerinae (15-20 Mya) (Ramírez et al. 2007) and the genera *Dendrobium* and *Earina* (20-23 Mya) (Conran et al. 2009) as the minimum ages for the respective clades. In addition, fossils of the oldest known Asparagales (93-105 Mya) (Iles et al. 2015) and monocot (110-120 Mya) (Iles et al. 2015) were used as minimum and maximum ages at the root of the tree. Two partitions, *matK* and *rbcL* were defined, following GTR+Γ (*matK*) and GTR+ Γ +I (*rbcL*) nucleotide substitution models, with an uncorrelated relaxed molecular clock (Drummond et al. 2006) assuming a log normal distribution of rates and a relaxed Yule speciation process.

Calibration points were modelled as follows: Goodyerinae [log normal distribution, mean 1.0, standard deviation 1.25, offset 15.0], *Dendrobium* [lognormal distribution, mean 1.0, standard deviation 1.25, offset 20.0], *Earina* [lognormal distribution, mean 1.0, standard

deviation 1.25] and the root [normal distribution, mean 106.5, standard deviation 8.21] following Guo et al. (2012).

The cloned *Xdh* dataset for estimating divergence times in Cypripedioideae comprised of sequences from Table 3.3.3 (214 samples), other *Paphiopedilum* subgroups analysed in section 2.3.3 and *Mexipedium*, *Phragmipedium*, *Selenipedium* and *Cypripedium* outgroups (see Table 2.1). A single partition was defined with an GTR+ Γ +I nucleotide substitution model following an uncorrelated relaxed molecular clock assuming a log normal distribution of rates and a Birth-Death speciation model. The crown age of *Paphiopedilum* was constrained using the result from the previous dating [normal distribution, mean 16.0, standard deviation 1.0].

All divergence estimations were conducted using a Bayesian inference approach implemented in the package BEAST v.1.8.0 (Drummond et al. 2012) on the CIPRES Science Gateway (Miller et al. 2010) (https://www.phylo.org/portal2) to infer a temporal framework for both datasets. Multiple runs of at least 40×10^6 generations were performed, sampling one tree every 1000th generation. Each parameter estimation was checked with Tracer v1.6 (Rambaut et al. 2013) to confirm effective sampling sizes had been achieved for each parameter and that runs had converged. Average branch lengths and 95% confidence intervals on nodes were calculated using TreeAnnotator v1.7.5 after burn-in and reported on maximum clade credibility consensus trees.

3.3.3 Geographic regions

Geographic regions discussed in this study are defined as (**A**) North Indochina, South China & the Himalayas, (**B**) South Indochina, (**C**) the Malaysian Peninsula, (**D**) Sumatra & Java, (**E**) Borneo, (**F**) the Philippines, (**G**) Wallacea and (**H**) Papua New Guinea (see Fig. 2.4 and).

3.3.4 Biogeoraphical inferences

I used the dispersal-extinction-cladogenesis (DEC) likelihood implemented in Lagrange v. 2.0.1 (Ree et al. 2005; Ree and Smith 2008) to infer the biogeographical history of *Paphiopedilum* section *Barbata* on an all-compat Bayesian tree produced from the concatenated plastid, *Xdh* and *CHS* loci dataset in Chapter 2. This dataset had previously been pruned to (1) remove samples suspected of undergoing recent hybridisation, defined as conflicting gene tree placements and (2) purely geographic groupings which defied morphological expectations, as means to elucidate the major historical dispersal patterns of *Paphiopedilum* section *Barbata*. The initial python script was generated using the Lagrange configurator available at http://www.reelab.net/lagrange/configurator/index. Details of the species distributions and

dispersal constraints are shown in Table 3.3 and 3.4. The biogeographical scenario is summarised as pie charts.

3.4 Results

3.4.1 Molecular dating

Molecular dating estimates for divergence times in section *Barbata* were obtained in two steps due to the lack of fossils and sequence data. First, an initial round of molecular dating was conducted on a plastid matrix comprised of *rbcL* and *matK* sequences for Orchidaceae calibrated with four fossils (Goodyerinae; Earinae; *Dendrobium* and Asparagales) which estimated the median crown age for the genus *Paphiopedilum* at around 16 My (Fig. 3.3). This date was used as a secondary calibration point on the Cypripedioideae *Xdh* dataset (Fig. 3.4). Using this approach, the median crown age of section *Barbata* was estimated at around 6.3 Mya with a 95% HPD range from 4.0 to 8.8 Mya. This estimate agrees with the recent plastid-based dates reported by Guo et al. (2015). A second diversification, which gave rise to two sister clades B1 and B2 (see below), is estimated to have occurred at around 5.1 Mya with a 95% HPD range from 3.4 to 7.2 Mya.

3.4.2 Analysis of the cloned *Xdh* sequences

In order to identify paralogues and allelic variants of *Xdh* within section *Barbata*, I cloned and Sanger sequenced samples that showed evidence of sequence ambiguities in the electropherograms from Chapter 2. In each case ten clones per plant were sequenced (Table 3.3). The cloning regime revealed a considerable diversity in the number of *Xdh* sequence variants. Comparatively, higher numbers of *Xdh* variants (up to at least eight) were found in samples from the *P. barbatum-callosum-lawrenceanum* and the *P. appletonianum-bullenianum* alliances (Table 3.3).

Analysis of the cloned *Xdh* sequences improved the phylogenetic resolution within section *Barbata* and resolved three well supported clades, Clade A, B1 and B2 (Fig. 3.5). Overall, there appeared to be greater phylogenetic resolution in Clade B1 relative to Clade B2 i.e. Clade B1 contained several subclades whereas B2 collapsed into a polytomy. This might be a reflection of the evolutionary ages of the clades and suggests that Clade B2 might be comparatively younger than Clade B1.

Clade A is comprised of just two taxa, *P. hookerae* (2*n*=28) and *P. sangii* (2*n*=28) from Borneo and Wallacea (Sulawesi) and is sister to the rest of section *Barbata*. The remaining two clades are primarily geographical. Clade B1 is composed predominantly of taxa with continental (North and South Indochina and the Malaysian Peninsula) distributions, namely *P. purpuratum* and members of the *P. sukhakulii-venustum-wardii* and *P. barbatum-callosum-lawrenceanum* alliances. Clade B2 on the other hand is composed mainly of taxa with an island (Borneo, Sumatra, Java, Wallacea, Philippines and Papua New Guinea) distributions, namely *P. tonsum*, *P. javanicum*, *P. superbiens*, *P. papuanum*, *P. violascens*, *P. argus*, *P. dayanum*, *P. hennisianum*, *P. sugiyamanum*, *P. inamorii*, *P. hennisianum* and *P. fowliei*.

The *P. barbatum-callosum-lawrenceanum* and *P. appletonianum-bullenianum* alliances which have wide distributions and occur on both continental and island SE Asia displayed a remarkable mixture of phylogenetic signals. Neither alliance was monophyletic (Fig. 3.5A-B & Table 3.3) with individuals from South Indochina, the Malaysian Peninsula and Borneo resolving in both Clades B1 and B2.

Some taxa showed conflicting phylogenetic signals between the *Xdh* copy types that originated from the same sample (Fig. 3.5A-B and Table 3.3). For example, for *P. lawrenceanum* D from Borneo, sequences D2 and D4 resolved in Clade B1 whereas sequences D1 and D3 resolved in Clade B2. Similar conflicts were also apparent in *P. callosum* A and B from the Malaysian Peninsula, *P. callosum* E from Indochina, *P. bullenianum* A and D from the Malaysian Peninsula and *P. bullenianum* G from Wallacea. The aforementioned patterns of mixed and conflicting signals from the analyses of cloned *Xdh* copies are likely caused by historical hybridisation events. These support the findings from Chapter 2 and point to recent gene-flow or dispersal between continental SE Asia, Borneo and Wallacea.

3.4.3 Biogeographical inferences from Lagrange

The geographical range evolution of *Paphiopedilum* section *Barbata* was extrapolated with Lagrange (Ree et al. 2005; Ree and Smith 2008) using the all-compatible Bayesian tree (Fig. 3.7) produced from the concatenated plastid, *Xdh* and *CHS* sequence dataset. The results of the analysis are shown on Fig. 3.8 and summarised in 3.9 and Table 3.3.

The probable ancestral range of *Paphiopedilum* section *Barbata* was reconstructed as North Indochina and Borneo, which are non-adjacent regions. The overall dispersal pattern was from a west to east fashion, originating from continental Asia into the Sunda and from there into the Sahul shelf. Reverse dispersals were recovered within the Sunda, but not on the Sahul region or between Sunda and Sahul. Borneo appears to be a stepping stone for dispersal into Sumatra and Java and the Philippines. Colonisation of the Sahul shelf was reconstructed

via Borneo and the Philippines through Wallacea into Papua New Guinea, although without support. Reverse dispersals were detected between Sunda shelf landmasses and the Philippines but not with Wallacea. Despite the proximity of the landmasses, no dispersals were detected between the Malaysian Peninsula and Sumatra and Java or Wallacea and Sumatra and Java.

3.5 Discussion

Like all Orchidaceae, members of section *Barbata* produce thousands of fine wind-dispersed seeds, which suggest they may be able to disperse across vast distances, and are likely capable of traversing across short stretches of sea, as evidenced by the occurrence of section *Barbata* on oceanic islands of Wallacea, the Philippines and on Papua New Guinea. While the possibility of dispersal across wide stretches of sea cannot be discounted, there is evidence that biogeographic vicariance (Guo et al. 2012) may be the predominant factor accounting for the present day distribution for all Cypripedioideae.

3.5.1 Molecular dating and biogeography

Accurate molecular dating of this group is inherently difficult due to the paucity of good orchid fossils for calibration. Molecular date estimates must always be interpreted cautiously, even more so in this case where the dates are inferred indirectly from a secondary calibration. Although the use of secondary calibrations runs the risk of potentially amplifying errors from the primary dating (e.g. Graur & Martin 2004), until such a time when more reliably dated fossils become available, the molecular dates presented here provide a useful basic time frame for studying the biogeographic history for this group.

The estimated crown age of *Paphiopedilum* section *Barbata* of c. 6.3 My inferred in this study places the origin of the group around the late Miocene and early Pliocene in a biogeographical period when (1) proto-southeast Asia had already formed and all the major landmasses were already in a layout similar to the modern day configuration (Fig. 3.6), (2) the prevailing climate was conducive to tropical wet forests (Fig. 3.2), and (3) the glacial and inter-glacial induced cycles of sea-level fluctuations had already commenced (Fig. 3.1). This coincides with the crown ages of many angiosperm lineages in the region (reviewed in Crayn et al. 2015), suggesting that the conditions during the late Miocene and early Pliocene drove a floristic boom in SE Asia.

Falling sea-levels would have been accompanied by an expansion in size of the SE Asian land masses. This expansion would have been especially dramatic on the Sunda shelf,

resulting in land connections between Borneo, Sumatra and Java with continental SE Asia during each glacial maximum. The climatic conditions and the exact coverage of suitable habitat available for colonisation on the exposed Sunda shelf (Sundaland) are not known although reconstructions of forest coverage during the last glacial maximum present a variety of scenarios (see Cannon et al. 2009).

The Lagrange reconstruction indicates that Wallacea was likely colonised via Borneo and the Philippines rather than Sumatra and Java. The absence of contributions from Java may be due to the sparsity of wet forest habitat on Java (Fig. 3.2). Clade A in the *Xdh* (Fig. 3.5) and concatenated (Fig. 3.7) phylogenetic reconstructions is sister to the rest of *Paphiopedilum* section *Barbata* and is found on Borneo and Wallacea (Sulawesi), and the widely distributed *P. bullenianum*, are found across the hypothetical Wallace line which approximates the border between the Sunda and Sahul plates and is characterised by a sharp transition in biota (chiefly fauna) of Asian and Australian origin. Sulawesi has a composite origin. SW Sulawesi was previously contiguous with the Sunda shelf, whereas the rest of the island either originated from the Sahul shelf or from volcanic subduction (Spakman and Hall 2010). The current distribution of Clade A and *P. bullenianum* is thus most likely to be due to a recent west-east dispersal across the deep water of the Makassar Strait although the estimated age of the clade and recent studies reviewed in Stelbrink et al. (2012) suggest that a vicariant origin is not impossible.

The remaining clades, B1 and B2, include species distributed in particular geographic regions. Clade B1 is composed primarily of continental samples (N & S Indochina and the Malaysian Peninsula), whereas Clade B2 contains mainly samples of island (Sumatra, Java, Wallacea, the Philippines and Papua New Guinea) origin, with Borneo sharing affinities with both regions. The diverse origins of Bornean taxa are supported by the dispersal patterns in the Lagrange analysis (Fig. 3.8). As found in Chapter 2, taxa with wide distributions, e.g. the *P. appletonianum-bullenianum* alliance, were found to be paraphyletic, with samples resolving into multiple clades according to their geographic origin (syngameons) rather than what their taxonomic affinity would suggest. The formation of syngameons may be caused by relatively higher incidence of hybridisation between sympatric taxa which suggests that interspecific barriers to gene flow are weak.

The Lagrange ancestral range reconstruction and the greater phylogenetic depth of Clade A and B1 suggest that these taxa of *Paphiopedilum* section *Barbata* are probably of continental or Bornean origin and that the taxa of Clade B2 are a result of a recent dispersal and rapid radiation on the SE Asian islands. The putative origin on continental SE Asia and

Borneo is surprising given the disjunction between the landmasses. A possible explanation for this result may be that the common ancestor of *Paphiopedilum* section *Barbata* might have once have had a wide distribution that encompassed continental SE Asia and Borneo but subsequently became extinct all other parts of its range.

The Lagrange reconstruction and *Xdh* hybridisation signals point to comparatively a higher frequency of exchanges between the mainland (Malaysian Peninsula and South Indochina) and Borneo than between the other Sunda landmasses. This phylogenetic pattern could be explained by (1) a higher connectivity (wet forest corridors) between Borneo and continental SE Asia that facilitated floristic exchange, and/or (2) the presence of old refugia forests in Borneo and continental SE Asia that captured and preserved the genetic diversity from Sundaland. The latter theory is supported by other studies (Cannon et al. 2009) which indicate the consistent presence of wet forests in Borneo throughout the last glacial maximum.

The Malaysian Peninsula, Borneo, Sumatra and Java share biotic affinities. This close biogeographic relationship is reflected in the sharing of taxa of section *Barbata*, i.e. *P. javanicum* is found on Borneo, Sumatra and Java. However, in contrast to expectation, there was no evidence of recent connectivity between Borneo and Sumatra in the *Xdh* dataset. However, connectivity is apparent in the plastid dataset presented by Guo *et al*. (2015) between the *P. lawrenceanum* and *P. tonsum* samples. Thus, the absence of a connectivity signal may simply be a sampling artefact. Analysis of additional loci will help to clarify this point.

3.6 Conclusions

Analysis of the cloned *Xdh* copies resolved *Barbata* into three distinct primary clades. The first, Clade A is sister to the rest of section *Barbata* and is composed of two species with a chromosome number (2*n*=28) unique to the section (but found elsewhere in the genus). The remaining two, Clades B1 and B2 are geographical groupings. Clade B1 contains mainly taxa of continental (North and South Indochina, the Malaysian Peninsula) origin whereas Clade B2 is composed primarily of island (Sumatra, Java, Wallacea, the Philippines and Papua New Guinea) taxa. Taxa from Borneo showed mixed phylogenetic affinities and appeared in both Clades B1 and B2. Samples of section *Barbata* from South Indochina, the Malaysian Peninsula and especially Borneo showed evidence of mixed phylogenetic signals i.e. multiple paralogous copies of *Xdh*, possibly as a result of historical hybridisation, supporting the idea of greater interconnectivity and/or the presence of refugia in those regions.

The estimated molecular age, Lagrange reconstruction and overall phylogenetic patterns suggest that section *Barbata* originated in Sundaland and has recently undergone a radiation on the islands. Colonisation of the Sunda landmasses was probably via a mixture of across-sea and shorter land-based seed dispersal whereas in the Philippines, Wallacea and Papua New Guinea, colonisation was most likely achieved exclusively by seed dispersal across short stretches of sea. Colonisation was likely to have been facilitated by receding sea-levels which exposed land bridges and narrowed the distances between landmasses. Borneo appears to be stepping stone for the exchanges between the Sunda landmasses. Wallacea was colonised from Philippines and Borneo but not Sumatra and Java.

Tables

Table 3.1 Geographic origins of plant material (A-F, as defined in Fig. 2.4) and number of *Xdh* copy types found in each sample. Also shown are the phylogenetic clades (see Fig. 3.5) in which the *Xdh* copy types are found. Highlighted rows show those species where *Xdh* clones fall in more than one clade, potentially indicating ancestral interspecific hybridisation.

Sample	Geographic origin	No. of *Xdh* copy types	Clade A	Clade B1	Clade B2
P. acmodontum A	F	1	0	0	1
P. acmodontum B	F	1	0	0	1
P. appletonianum A	B	4	0	4	0
P. appletonianum B	B	7	0	7	0
P. appletonianum C	A	1	0	1	0
P. appletonianum D	B	4	0	4	0
P. appletonianum E	B	2	0	2	0
P. appletonianum F	B	6	0	6	0
P. appletonianum G	B	1	0	1	0
P. appletonianum H	A	1	0	1	0
P. appletonianum I	A	1	0	1	0
P. argus A	F	4	0	0	4
P. argus B	F	2	0	0	2
P. barbatum A	C	1	0	1	0
P. barbatum B	C	1	0	1	0
P. barbatum C	C	1	0	1	0
P. barbatum E	C	2	0	2	0
P. barbatum F	C	1	0	1	0
P. barbatum G	C	1	0	1	0
P. barbatum H	C	1	0	1	0
P. barbatum I	C	1	0	1	0
P. bullenianum A	C	5	0	1	4
P. bullenianum C	E	1	0	1	0
P. bullenianum D	C	1	0	0	1
P. bullenianum F	D	3	0	0	3
P. bullenianum G	G	2	0	1	1
P. bullenianum H	C	1	0	1	0
P. bullenianum J	G	1	0	1	0
P. bullenianum K	C	1	0	1	0
P. bullenianum L	E	7	0	7	0
P. bullenianum M	C	1	0	1	0
P. bullenianum N	E	8	0	8	0
P. callosum A	C	8	0	3	5
P. callosum B	C	3	0	1	2
P. callosum C	C	1	0	0	0
P. callosum D	B	1	0	0	0
P. callosum E	B	7	0	6	1

(cont.)

Table 3.1 (cont.)

Sample	Geographic origin	No. of *Xdh* copy types	Clade A	Clade B1	Clade B2
P. callosum F	B	1	0	1	0
P. callosum G	B	1	0	1	0
P. ciliolare	F	1	0	0	1
P. dayanum A	E	3	0	0	1
P. dayanum B	E	1	0	0	1
P. fowliei A	F	1	0	0	1
P. fowliei B	F	2	0	0	2
P. hennisianum A	F	1	0	0	1
P. hennisianum B	F	1	0	0	1
P. hookerae A	E	1	1	0	0
P. hookerae B	E	1	1	0	0
P. hookerae C	E	2	2	0	0
P. hookerae D	E	1	1	0	0
P. inamorii	E	1	0	0	1
P. javanicum A	D	1	0	0	1
P. javanicum B	D	1	0	0	1
P. javanicum C	D	1	0	0	1
P. javanicum D	E	1	0	0	1
P. javanicum E	D	1	0	0	1
P. javanicum F	D	1	0	0	1
P. lawrenceanum A	E	1	0	1	0
P. lawrenceanum B	E	1	0	1	0
P. lawrenceanum C	E	1	0	1	0
P. lawrenceanum D	E	4	0	2	2
P. lawrenceanum E	E	1	0	1	0
P. lawrenceanum F	E	1	0	1	0
P. mastersianum A	G	4	0	0	4
P. mastersianum B	G	1	0	0	1
P. mastersianum C	G	1	0	0	1
P. mastersianum E	G	2	0	0	2
P. papuanum A	H	2	0	0	2
P. papuanum B	H	1	0	0	1
P. purpuratum	A	3	0	3	0
P. sangii A	E	1	1	0	0
P. sangii B	E	2	2	0	0
P. sangii C	E	2	2	0	0
P. schoseri	G	1	0	0	1
P. sugiyamanum	E	1	0	0	1
P. sukhakulii A	B	3	0	3	0
P. superbiens A	C	1	0	0	1
P. superbiens B	C	1	0	0	1
P. superbiens C	C	1	0	0	1

(cont.)

Table 3.1 (cont.)

Sample	Geographic origin	No. of *Xdh* copy types	Clade A	Clade B1	Clade B2
P. superbiens D	C	4	0	0	4
P. superbiens E	C	1	0	0	1
P. superbiens G	C	1	0	0	1
P. tonsum A	C	1	0	0	1
P. tonsum B	C	1	0	0	1
P. tonsum C	C	1	0	0	1
P. tonsum D	C	1	0	0	1
P. tonsum E	C	1	0	0	1
P. tonsum F	C	1	0	0	1
P. urbanianum A	F	1	0	0	1
P. urbanianum B	F	2	0	0	2
P. venustum A	A	1	0	1	0
P. venustum B	A	1	0	1	0
P. venustum C	A	1	0	1	0
P. venustum D	A	1	0	1	0
P. violascens A	H	1	0	0	1
P. violascens B	H	1	0	0	1
P. violascens C	G	1	0	0	1
P. wardii A	A	5	0	5	0
P. wardii B	A	4	0	5	0
P. ×siamense	B	2	0	2	0

Table 3.2 Dispersal constraints used in the Lagrange analysis

	A	B	C	D	E	F	G	H
A	-	1.0	0.01	0.01	0.01	0.5	0.01	0.01
B	1.0	-	1.0	0.01	0.5	0.5	0.01	0.01
C	0.01	1.0	-	1.0	0.5	0.01	0.01	0.01
D	0.01	0.01	1.0	-	0.5	0.01	0.01	0.01
E	0.01	0.5	0.5	0.5	-	0.5	0.5	0.01
F	0.5	0.5	0.01	0.01	0.5	-	0.5	0.01
G	0.01	0.01	0.01	0.01	0.5	0.5	-	0.5
H	0.01	0.01	0.01	0.01	0.01	0.01	0.5	-

Table 3.3 Biogeographic events inferred by the Lagrange analysis in Fig. 3.8

Clade Posterior Probabilities	Node/Tip	Event
1	96	**V** E|A
1	96 to 97	**D** AB
0.04	106 to 135	**D** EC
1	107 to 108	**D** ED
1	107	**V** D|E
1	107 to *P. schoseri*	**D** EG + **E**E
0.13	113 to 114	**D** DE
0.84	114	**V** E|D
0.84	114 to 115	**D** EF
0.85	115	**P** F|EF
0.85	115 to 125	**E**E + **D** FG
0.07	116 to 120	**D** FE
0.02	120	**P** F|EF
0.07	125	**V** G|F
0.07	125 to 126	**D** GH
0.86	126	**V** H|G
0.27	97	**P** B|AB
0.12	98 to 101	**D** BE
1	101	**V** B|E
0.98	102 to 103	**D** BC
0.97	103	**V** C|B
0.92	99 to 100	**D** BA
0.57	90 to 91	**D** EC
0.67	91 to 93	**D** EG

Event Key:
D – dispersal
E – extinction
P – peripatric speciation
V – vicariance

Figures

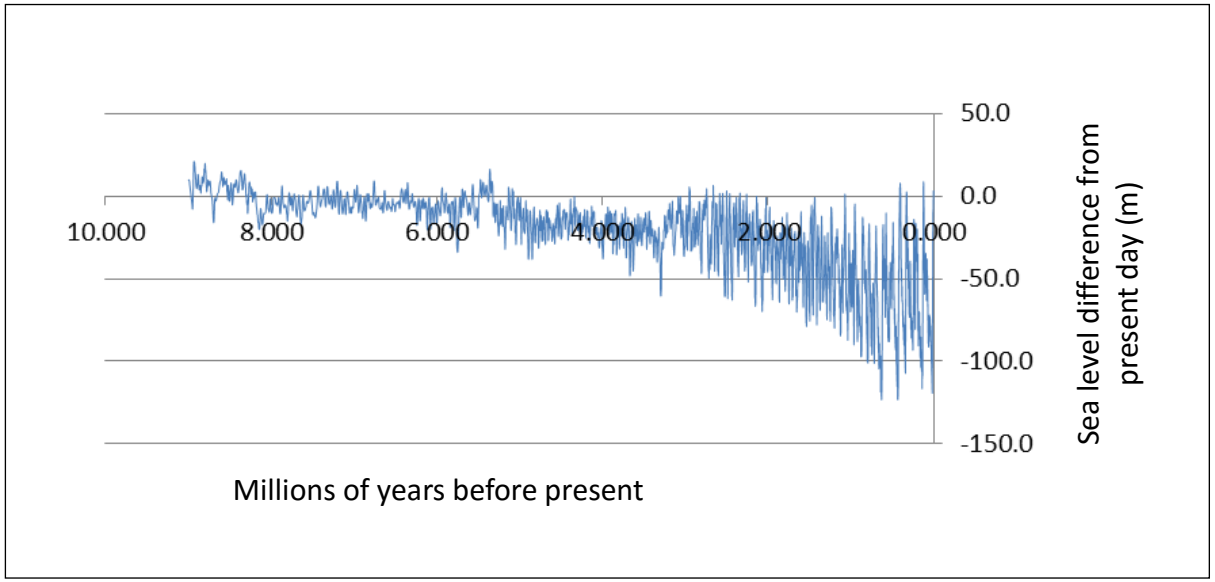

Figure 3.1 Sea level (m) changes calculated from oxygen isotopes (adapted from Miller et al. 2011).

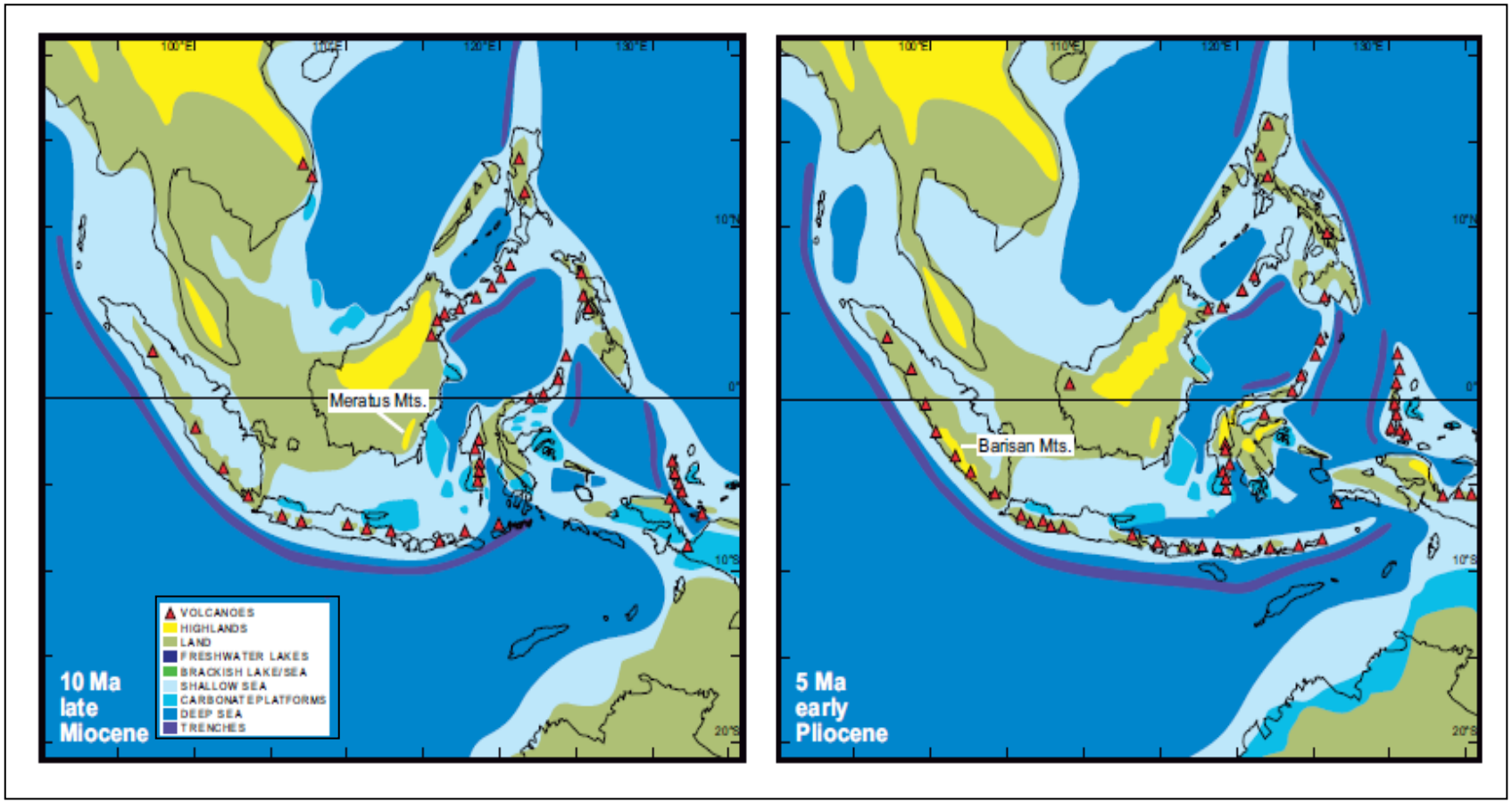

Figure 3.2 The proposed distributions of land, lakes and sea in South East Asia 5 and 10 million years ago, showing the predicted extent of tropical forest cover (green) (reproduced from de Bruyn et al. 2014).

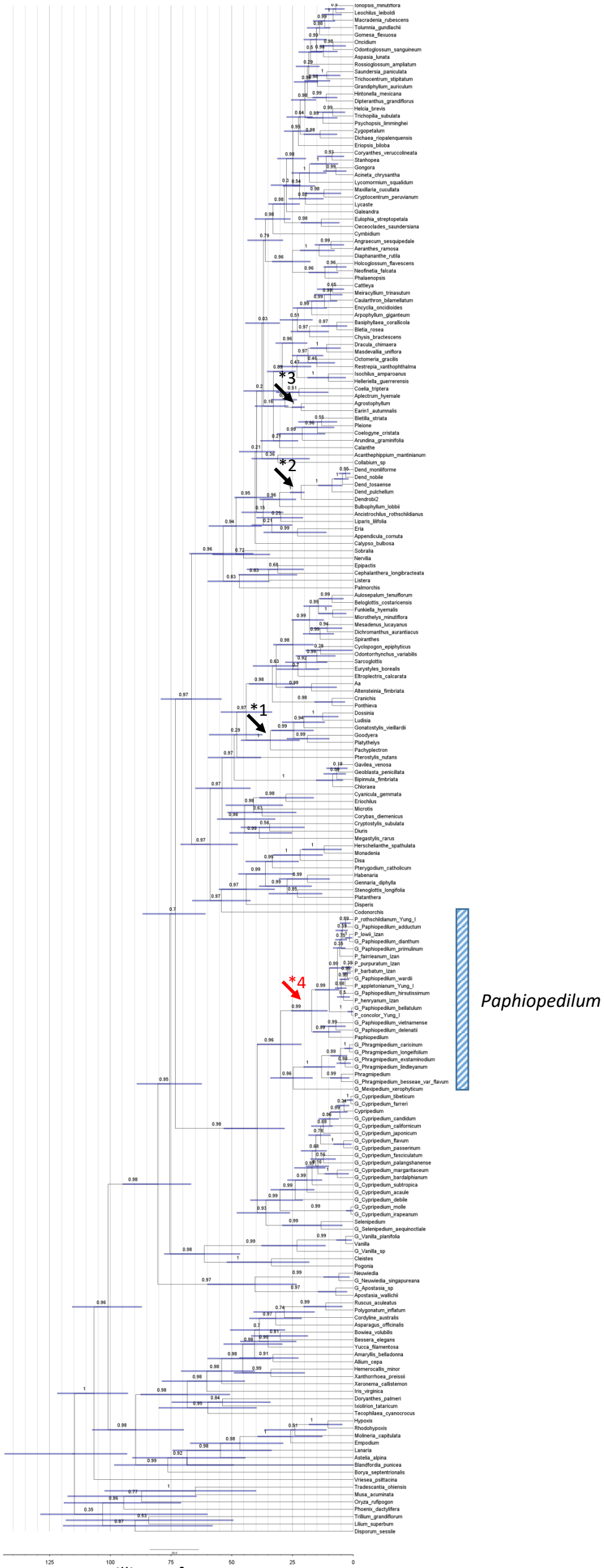

Figure 3.3 A dated phylogenetic tree of Orchidaceae produced from BEAST analysis of plastid *matK* and *rbcL*. The tree has been calibrated as minimum ages for the respective clades, using orchid fossils of: *1 Goodyerinae (15-20 Mya) (Ramírez et al. 2007); *2 *Dendrobium* (20-23 Mya) and *3 *Earina* (20-23 Mya) (Conran et al. 2009). Bars represent node ages (95% HPD). Numbers above branches indicate posterior probabilities (PP). The clade comprising species of *Paphiopedilum* (*4, in red) has a median age of c. 16 million years.

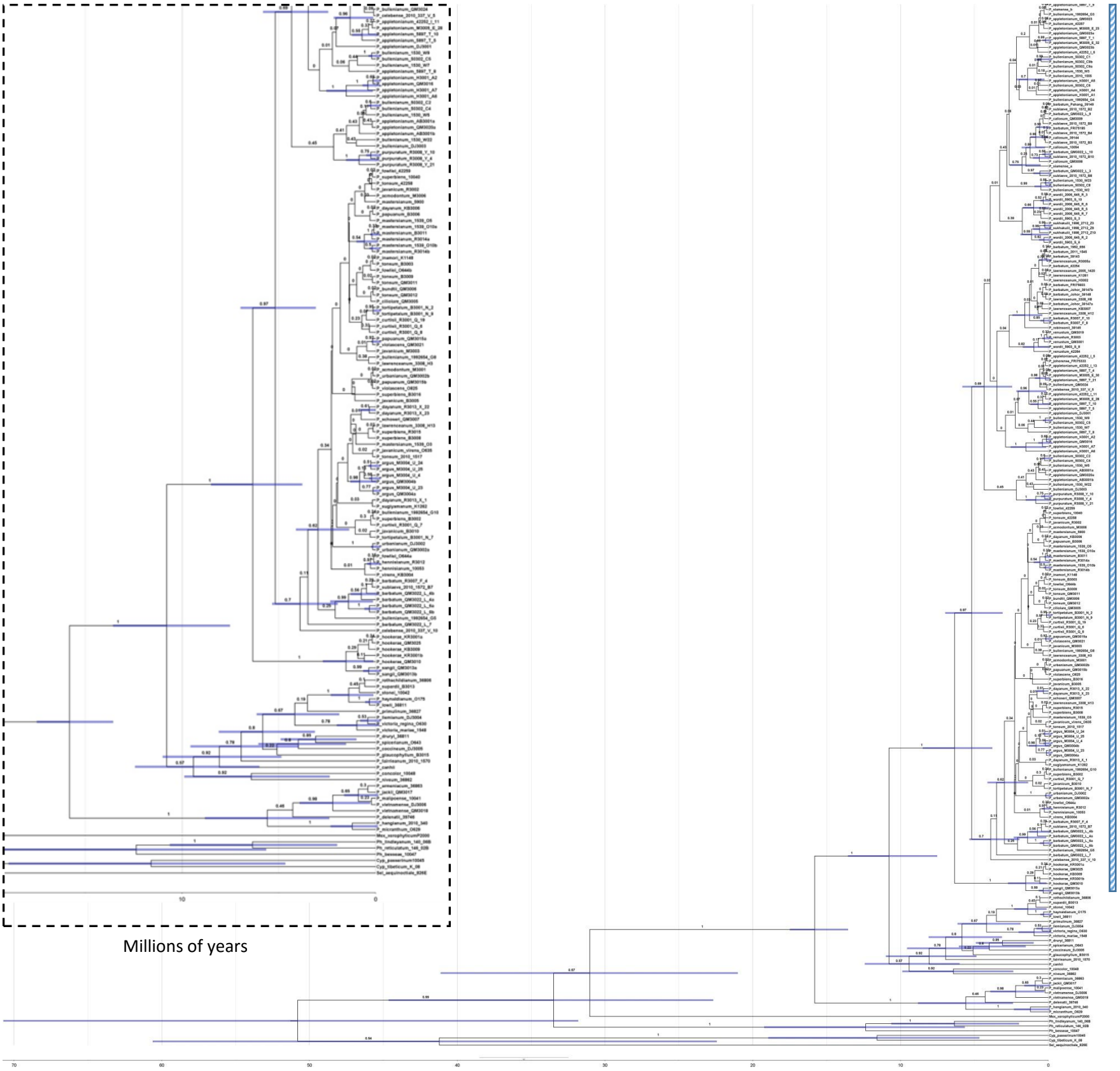

Figure 3.4 A dated phylogenetic tree of *Paphiopedilum* produced from BEAST analysis of cloned nuclear *Xdh* constrained with a secondary calibration derived point from Fig. 3.3. Numbers above branches represent posterior probabilities (PP) support. Bars indicate node ages (95% HPD) for clades with >0.5 PP. Because of the size of the tree the resolution of the image presented is reduced, so an expanded area is shown in the insert (top left).

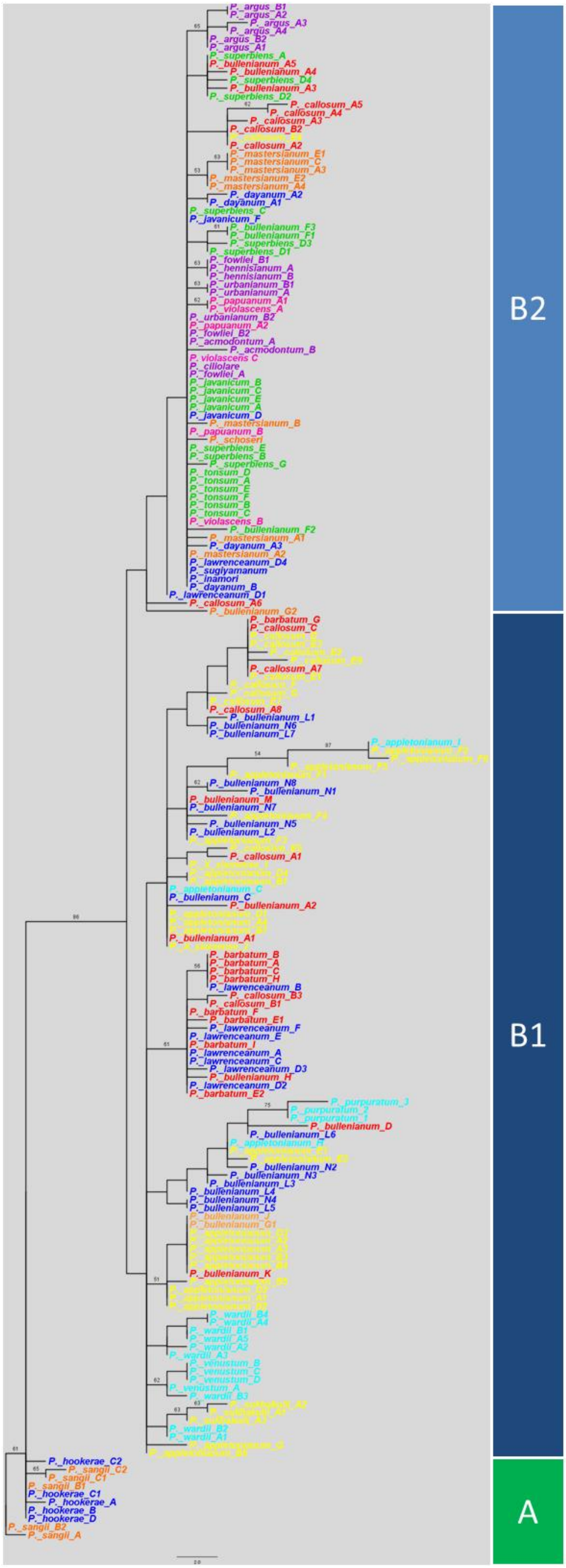

Figure 3.5A A tree showing the taxon relationships within *Paphiopedilum* section *Barbata* produced by MP analysis of the cloned nuclear *Xdh* sequences. Numbers above branches indicate bootstrap percentages (BP) above 50. The colours of the tips indicate their origin and correspond to the geographical ranges displayed in Fig. 2.4. Three clades A, B1 and B2 are indicated.

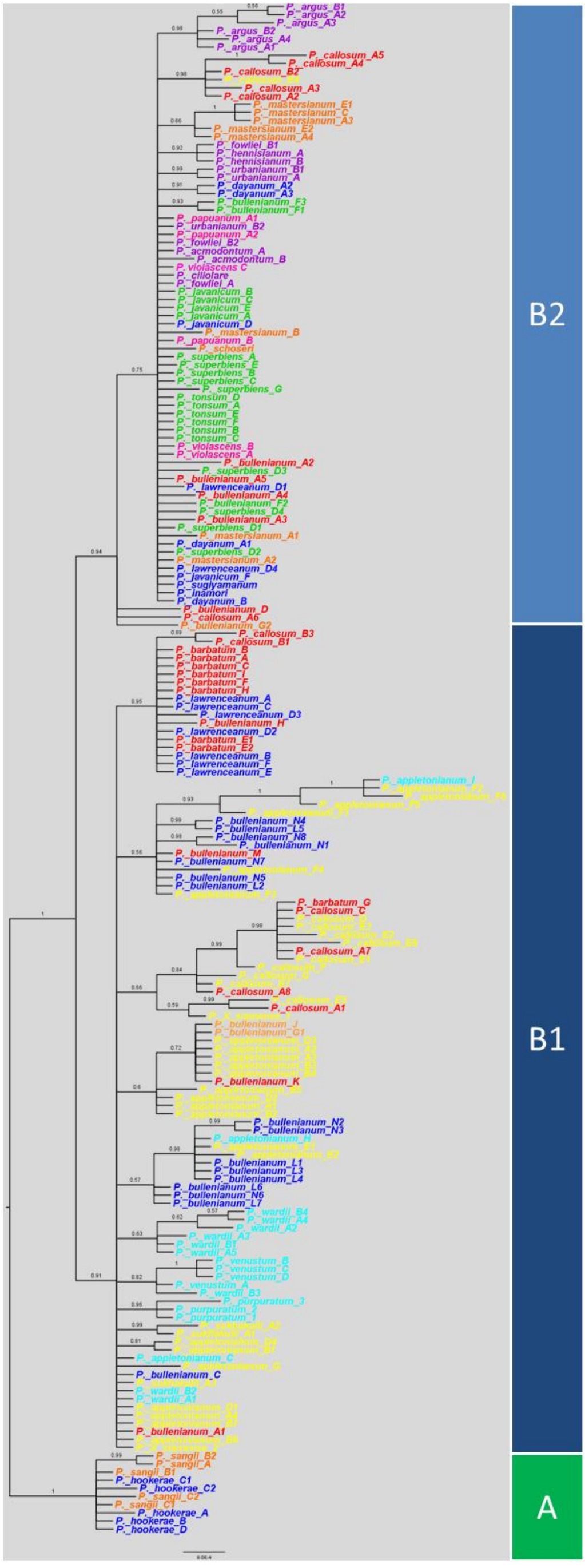

Figure 3.5B A tree showing the taxon relationships within *Paphiopedilum* section *Barbata* produced by Bayesian analysis of the cloned nuclear *Xdh* sequences. Numbers above branches indicate posterior probabilities (PP) above 0.5. The colours of the tips indicate their origin and correspond to the geographical ranges displayed in Fig. 2.4. Three clades A, B1 and B2 are indicated.

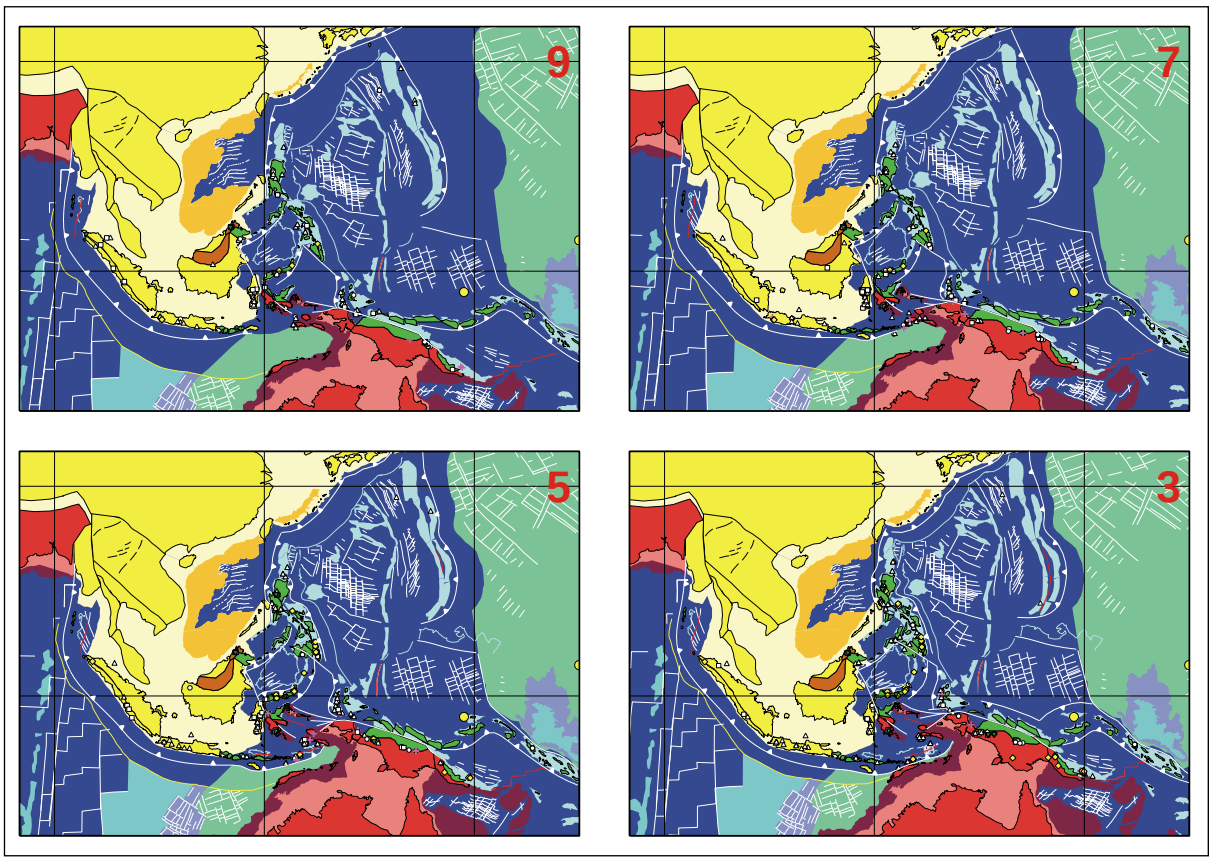

Figure 3.6 Plate tectonic reconstructions of SE Asia at 9, 7, 5 and 3 Mya (reproduced from Hall 2002) showing the locations of major landmasses.

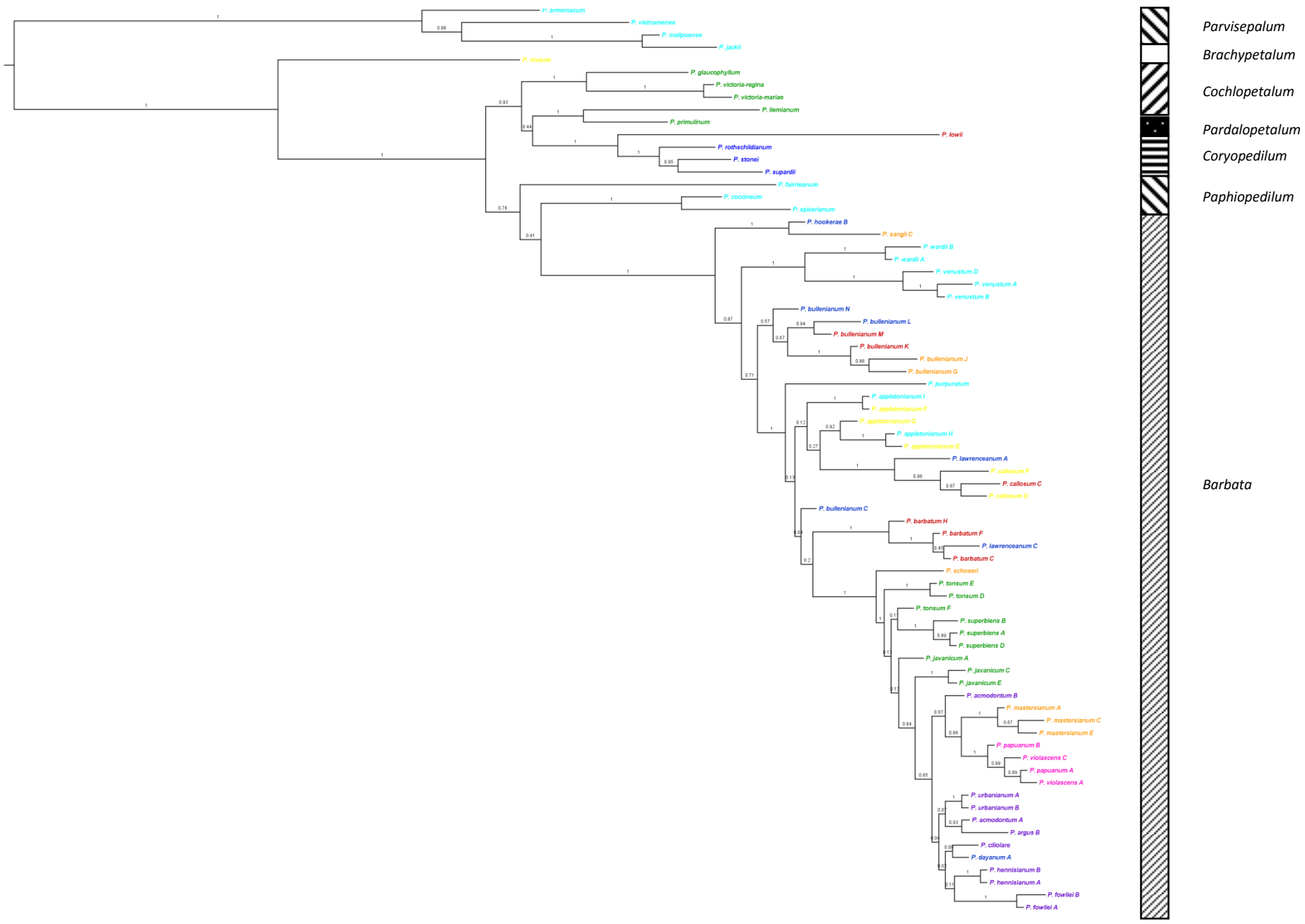

Figure 3.7 All-compatible consensus tree produced by Bayesian analysis of the concatenated combined plastid, nuclear *Xdh* and *CHS* regions. Numbers above branches indicate posterior probability (PP) scores above 0.5. Tip colours correspond to the geographic regions shown in Fig. 2.4.

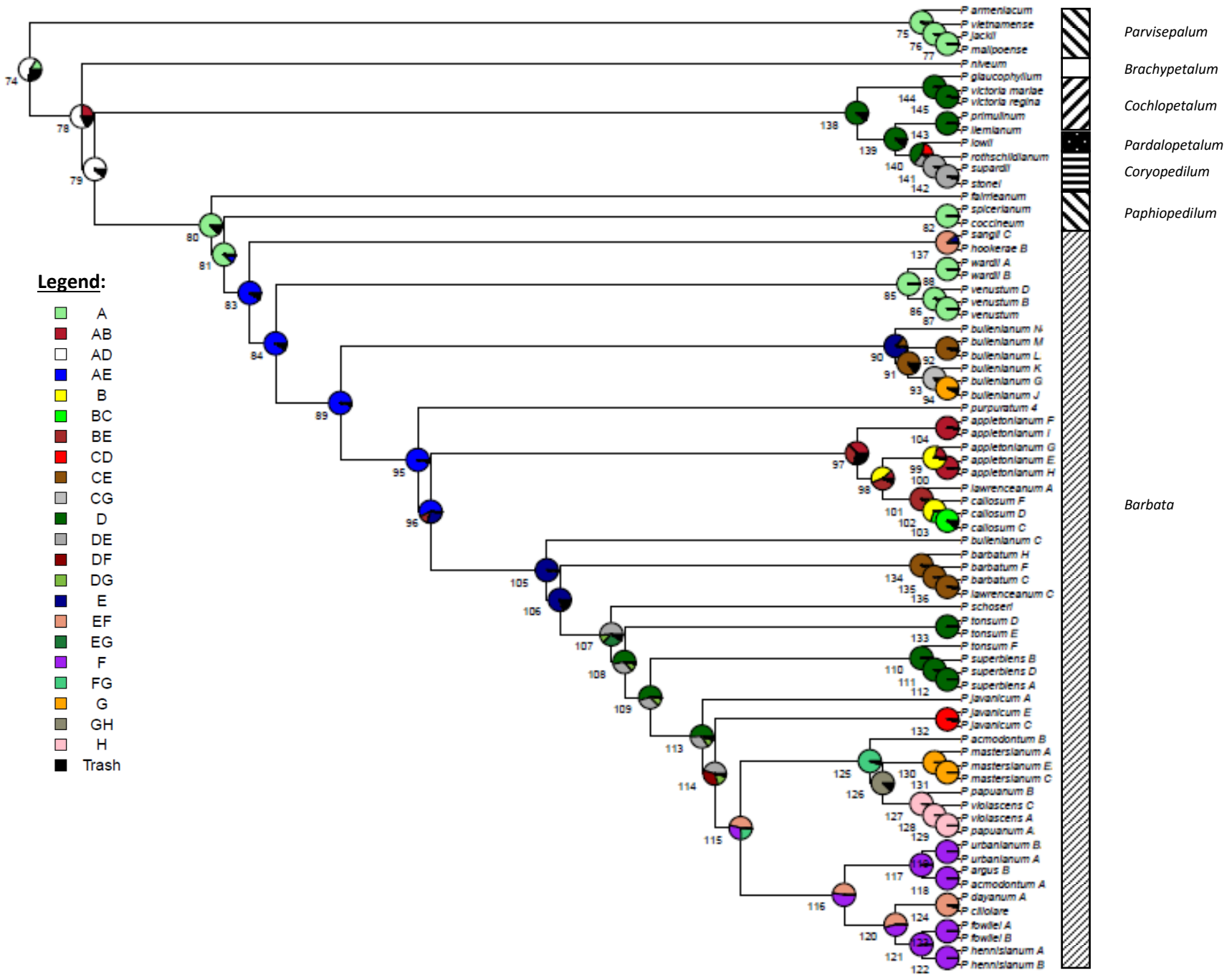

Figure 3.8 Biogeographical reconstruction inferred by the Lagrange analysis on the all-compatible consensus tree presented in Fig. 3.7. Biogeographical events along the tree nodes are summarised in Table 3.8

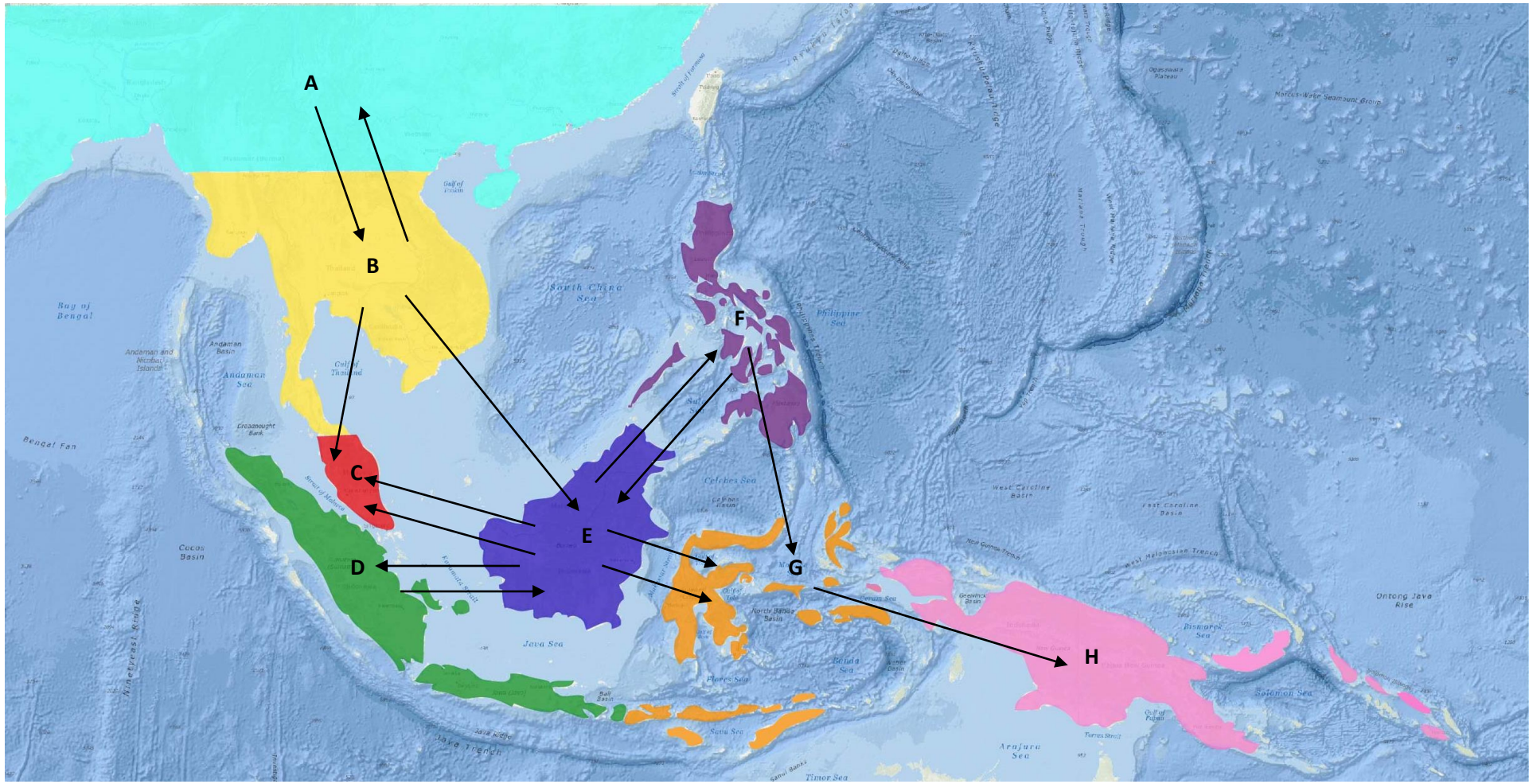

Figure 3.9 A map of SE Asia showing the biogeographic reconstruction of *Paphiopedilum* section *Barbata* inferred from the Lagrange analysis from Table 3.8. The arrows indicate dispersal events. Coloured areas are (**A**) North Indochina, South China & Himalayas, (**B**) South Indochina, (**C**) The Malaysian Peninsula, (**D**) Sumatra & Java, (**E**) Borneo, (**F**) The Philippines, (**G**) Wallacea, and (**H**) Papua New Guinea

CHAPTER 4
GENOME EVOLUTION

4.1 Abstract

The first flow cytometry (FCM) genome size estimates are presented for 25 *Paphiopedilum* taxa. *Paphiopedilum* genome sizes (2C) varied but were all large and exceeded 45 pg. Generally, higher genome size values were observed in section *Barbata* than in other *Paphiopedilum* groups. Within section *Barbata* (2*n*=28-42) genome sizes were found to range from 56.5 (*P. bullenianum*, 2*n*=40, 42) to 79.0 pg (*P. sangii*, 2*n*=28) and appeared to be independent of chromosome number. The genome size values are complemented by new high-throughput sequence data to provide insights into the repetitive fraction of the genome in seven taxa, representing six of the main *Paphiopedilum* subgenus/section lineages and a *Phragmipedium* outgroup. The sequence data were characterised with the Galaxy-based RepeatExplorer (RE) graph clustering program. This approach revealed that some 50.4 to 62.0% of the genomes in *Paphiopedilum* were comprised of repetitive DNA. The most abundant repeat families (comprising between 31.8 and 48.9% of the genomes) were *Ogre/Tat* transposable elements (TE), which belong to the *Ty3/gypsy* long terminal repeat (LTR) retrotransposons. However, there was no clear association between increasing genome size and the specific gain or loss of any one particular type of repetitive DNA. Instead, genome size increase in *Paphiopedilum* appeared to be caused by an overall increase in all types of repetitive DNA rather than any one family or type of element. In addition, *Paphiopedilum* genomes appeared to harbour a large proportion (28.9 to 39.5%) of single or low-copy DNA, possibly derived from ancient repetitive elements. Collectively these findings suggest that low rates of repetitive DNA removal rather than the proliferation of any particular family of repetitive element is driving genome evolution in the group, possibly mediated via epigenetic processes. The karyotype and genome size evolution in *Paphiopedilum* is re-evaluated in the light of the new FCM, RE clustering and phylogenetic data.

4.2 Introduction

In simple terms, the 2C genome size can be defined as the total amount of DNA in an unreplicated diploid nucleus. Genome size values are typically expressed in picograms (pg) (1 pg=978 Mbp) (Doležel et al. 2003). The ongoing collation of data on genome sizes in plants available through the Plant DNA C-values database (compiled in Bennett and Leitch 2012) has

revealed a remarkable diversity in genome sizes. In seed plants alone, an astounding c. 2400-fold variation in genome sizes has been detected, with 2C-values ranging from approximately 0.1 pg in *Genlisea tuberosa* (Lentibulariaceae) (Fleischmann et al. 2014) to 304.5 pg in *Paris japonica* (Melanthiaceae) (Pellicer et al. 2010). However, the distribution of genome sizes is highly asymmetrical with the overwhelming majority of seed plants possessing small genomes, whereas large (2C≥28 pg) and very large (>70 pg) genomes occur only in certain lineages (Leitch et al. 2010; Kelly and Leitch 2011). While the largest range in genome sizes of any family of seed plants is found in Melanthiaceae, with C-values ranging 230-fold (Pellicer et al. 2014), Orchidaceae have the second largest range with values ranging 168-fold (2C=0.6-110.8 pg) with the biggest genomes found in Cypripedioideae and Vanilloideae (Leitch et al. 2009).

Large genomes can arise by (1) whole genome duplication (polyploidy) through somatic chromosome doubling or, probably more frequently, via fertilisation of unreduced gametes (reviewed in Bennett and Leitch 2005; Hegarty and Hiscock 2008), as is especially common in some ferns with large genomes (e.g. *Ophioglossum reticulatum*, 2*n*=c. 1440; Abraham and Ninan 1954; Obermayer et al. 2002; Nakazato et al. 2008), or (2) polyploid-independent genome expansion, such as in *Fritillaria* (Liliaceae) which contains the largest known diploid species to date (Ambrozová et al. 2011). While the former mechanism of genome size expansion is relatively well understood, the evolutionary forces leading to polyploid-independent genome expansion and downsizing remains mysterious, although increases have been attributed to the accumulation of transposable elements (Vitte and Panaud 2005; Hawkins et al. 2006; Piegu et al. 2006) while decreases are considered to arise from recombination-based processes such as unequal and illegitimate recombination (Bennetzen 2005).

In order to increase understanding of the genomic processes operating in plants with large genomes, I analysed the genomes of *Paphiopedilum* species belonging to section *Barbata* (Cypripedioideae), an evolutionarily young (crown age=6.3 Mya, 95% HPD=4.0-8.8 Mya, see Chapter 3) orchid group which displays considerable genome size and chromosome number diversity (reviewed in Chapter 1). Illumina HiSeq sequences of representative samples of subgenera and sections of *Paphiopedilum* were obtained and their repetitive DNA content analysed with *RepeatExplorer* (RE), a bioinformatics pipeline using a Galaxy-based web server for genome-wide characterisation of repetitive elements. In addition, I extended the existing knowledge on genome size variation in *Paphiopedilum* by generating new genome size estimates, determined using flow cytometry (FCM), and revisited existing theories on karyotype and genome size evolution in the group in the light of the new data.

Low-coverage high-throughput sequencing is increasingly used for plastid and mitochondrial genome assemblies. These generate not only the target organellar DNA, but also sequences from the repetitive part of the genome. Although often ignored, the repetitive sequences represent a source of phylogenetic information. A recently described method (Dodsworth, Chase, Kelly, et al. 2015; Dodsworth, Chase, Särkinen, et al. 2015) has been shown to be useful for inferring lower-level relationships across several genera from RE clusters. I therefore decided to evaluate the efficacy of this method on *Paphiopedilum*, an evolutionarily young and diverse group in which reticulate evolution is suspected.

4.3 Materials & Methods

4.3.1 Plant materials

Genome size estimation – A total of 32 samples, which included multiple individuals per taxon, were used for genome size estimates. The samples included 13 taxa from section *Barbata*, one taxon from subgenus *Brachypetalum*, one taxon from section *Cochlopetalum*, two taxa from section *Coryopedilum*, two taxa from section *Pardalopetalum*, five taxa from section *Paphiopedilum*, two taxa from subgenus *Parvisepalum* and a *Phragmipedium* outgroup. All plant materials used for genome size estimation originated from living collections at the Royal Botanic Gardens, Kew, Queen Mary University of London, Hortus botanicus, Leiden, Netherlands and Munich Botanic Gardens, Germany. Details of these samples are given in Table 4.1.

RepeatExplorer clustering – Genomic DNA for the *P. primulinum* sample used in RE clustering in Section 4.3.3 and Section 4.3.4 came from the Royal Botanic Gardens, Kew, and was sequenced at the Genome Centre, Queen Mary University of London. Dr. Yung-I Lee, at the National Museum of Taiwan, provided HiSeq sequence data for *P. armeniacum*, *P. concolor*, *P. rothschildianum*, *P. villosum* and *P. appletonianum* and Dr. Rene Smulders, Wageningen UR, provided HiSeq sequences for *P. druryi*, *P. henryanum*, *P. lowii*, *P. glanduliferum* A & B, *P. barbatum*, *P. purpuratum* and *Phragmipedium longifolium* (see Table 4.2).

4.3.2 Estimating genome sizes with flow cytometry

I used flow cytometry (FCM) to estimate the genome sizes of 25 *Paphiopedilum* taxa with emphasis on section *Barbata*. The Ebihara nuclei isolation buffer contained 1.0% Triton-100, 140 mM 2-mercaptoethanol, 50 mM Na_2SO_3, 50 mM Tris-HCl [pH 7.5], 25 µg/mL propidium

iodide, 40 mg/mL polyvinyl-pyrrolidone [PVP-40] and 0.1 mg/mL ribonuclease (Ebihara et al. 2005). *Vicia faba* 'Inovec' 2C=26.9 pg (Doležel et al. 1998) was used as an internal reference standard. Samples were prepared as in Ebihara et al. (2005). Sample and reference standard leaves were co-chopped in 2.0-mL of Ebihara buffer on ice in a fume hood. The resulting slurry was incubated for 5 min on ice and filtered through a 30-µm nylon mesh (Partec) into a 2.0-mL tube. The filtrate was incubated for 15 min at 37˚C and then left on ice for approximately 30 min. Genome sizes were measured on a Partec CyFlow SL equipped with a 100 mW Cobalt Samba 532 nm green laser. Three pseudo-replicate measurements of 5000 nuclei were made for each sample.

4.3.3 Processing of Illumina HiSeq data

The quality of the raw-reads was assessed by FastQC v0.10.1. Raw-reads were trimmed and quality filtered with a FASTX-Toolkit v0.013 to give 90-bp length reads with 90% of bases having a minimum Phred score of 20.

In order to remove sequences of organellar origin, custom Perl scripts and the stand-alone version of BLAST v2.2.16 (Altschul et al. 1997) were used to rapidly screen the filtered-reads against a custom database containing the draft plastid genome of *Phragmipedium longifolium* (unpublished data provided by Dr. W. Mark Whitten, University of Florida, USA) and the published mitochondria genomes of ten monocots (taken from NCBI). Parameter settings used for BLASTN searches were: -v 1 -G 0 -E 2 -K 0 -b 0 - e 0.000001 -F mL. Sequences with similarity matches to the database (*E*-value ≤19×10^{-6}) were removed. All remaining reads were considered to be of nuclear origin and uploaded onto the Galaxy server environment (http://www.repeatexplorer.org) as described in Dodsworth et al. (2015).

Nuclear reads were paired using FASTQ interlacer tool implemented in RE. As overlapping read pairs, which are caused by the presence of very short genomic DNA fragments in samples, can adversely affect RE clustering analysis by masking the connectivity between clusters, they were removed using the RE utilities (minimum overlap=30 nt, maximum mismatch per 100 bp=1, offset=5).

4.3.4 Clustering and annotation of repetitive DNA with RepeatExplorer

A total of eight HiSeq samples, representing individuals from the seven major subgroups of *Paphiopedilum* (*Parvisepalum*, *Brachypetalum*, *Coryopedilum*, *Cochlopetalum*, *Paphiopedilum*, *Pardalopetalum* and *Barbata*) and a *Phragmipedium* outgroup, were selected for RE clustering

(Table 4.2). The highest possible number of reads (Table 4.2) was used in order to maximise clustering. Clustering of Illumina reads was performed using the RE pipeline, implemented in a Galaxy server environment to identify and quantify repetitive DNA in *Paphiopedilum*. Simply put, RE runs a graph-based clustering algorithm (Novák et al. 2010, 2013) to identify groups of frequently overlapping reads representing families of repetitive elements followed by similarity- and structure-based repeat identification tools that aid in identification. RE uses a BLAST threshold of 90% similarity over 55% of the read to identify clusters based on a principle of maximum modularity. The RE clustering tool was used to identify repeat clusters within each dataset (minimum overlap=55, cluster threshold=0.01%, minimum overlap for assembly=40).

Large repeat clusters composed of tandem repeats (e.g. satellites) are problematic as they consume high amounts of computing resources and thus reduce the number of sequences that can be handled by the RE pipeline. As preliminary runs indicated that the *P. armeniacum* sample contained a high (>10%) proportion of satellite DNA, it was subjected to a custom sequence filter (containing a database of 617 of the most abundant sequence contigs from *SatA*) to remove 90% of satellite sequences, as per the RE developer's recommendations, prior to clustering.

Clusters were identified manually as described in Novák et al. (2010) by scanning read similarity hits to the RepeatMasker (Smit et al. 1996) database, visual examination of graphs and the location of cluster mates as well as sequence searches with BLASTX and BLASTN (Altschul et al. 1990) on public databases. In addition, selected cluster graphs were also analysed with SeqGrapher (http://cran.rproject.org/web/packages/SeqGrapheR/index.html). Only clusters with genome proportions (GP) greater than 0.01% were included in the analysis.

4.3.5 Combined comparative clustering and phylogenetic analysis of repeats

A combined *Paphiopedilum* dataset comprised of reads from 15 samples representing seven of the major *Paphiopedilum* subgroups (*Parvisepalum*, *Brachypetalum*, *Coryopedilum*, *Cochlopetalum*, *Paphiopedilum*, *Pardalopetalum* and *Barbata*) and a *Phragmipedium* outgroup was prepared for combined clustering. Details of samples are given in Table 4.2. For each taxon, the number of reads used was in proportion to its genome size to account for the genome size variation across the group.

In order to utilise the clusters as continuously varying characters for maximum parsimony phylogenetic analysis, the top 100 most abundant RE clusters and their relative

abundances were selected and used to create a character matrix in TNT (Goloboff et al. 2008). As TNT requires all values to fall between 0 and 65, we divided all raw cluster values by 700 [factor=largest cluster abundance in dataset/65]. We followed the settings in Dodsworth, Chase, Kelly, et al. (2015) [characters=normal additive characters, count changes=non-integers]. Tree searches were performed using the branch and bound algorithm with 10,000 symmetric bootstrap replicates.

4.3.6 Phylogenetic analysis of ITS and plastid sequences

Nuclear ITS and plastid (*ycf1* and *matK*) sequences were obtained by the mapping of nuclear and organellar paired-reads against reference sequences from GenBank (Benson et al. 2013) using Geneious v.7.1.8 (Biomatters Ltd, Auckland, NZ) (Kearse et al. 2012). All sequences were examined and aligned by eye and entered into nuclear ITS and combined plastid matrices. Maximum parsimony analyses were conducted using the same settings given in in Chapter 2, section 2.3.4.

4.3.7 Fluorescence *in situ* hybridisation probe design

Sequence contigs of the most abundant satellite cluster *SatA* (Fig. 4.2) generated from the comparative analysis in section 4.3.4 were downloaded from RE and aligned by Clustal v2.1 on Geneious v7.1.8 (Biomatters Ltd, Auckland, NZ) (Kearse et al. 2012) software. A commercially 5'-end-digoxigenin-labelled synthetic oligonucleotide probe:

[AAATCTGACCTAATTTGGACCCAATCTTTGAACCTTCTAATTGAAGGTCAATTGGTGT],

was designed based on the consensus alignment of the most abundant *SatA* sequence contigs. Additional probes consisted of 45S ribosomal DNA (rDNA) [i.e. pTa71 containing a repetitive unit of the 45S rDNA from *Triticum aestivum* (Gerlach and Bedbrook, 1979)] and 5S rDNA (pTa794 containing the 5S rDNA coding region from *T. aestivum* (Gerlach and Dyer, 1980). Both rDNA probes were labelled by nick translation with either digoxigenin-11-dUTP (Roche Diagnostic GmbH, Penzberg, Germany) or biotin-16-dUTP. The FISH procedure was carried out by Dr. Yung-I Lee at the National Museum of Science, Taiwan using the protocol described in Lee et al. (2011). The digoxigenin-labelled probes were detected by anti-digoxigenin-rhodamine (Roche Diagnostics GmbH) and the biotin-labelled probes by fluorescein isothiocyanate (FITC)-conjugated avidin (Vector Laboratories, Burlingame, CA, USA). Chromosomes were counterstained with 4', 6-diamidino -2-phenylindole (DAPI) in an antifade solution (Vector Laboratories, CA, USA). All images were captured digitally by a CCD camera attached to an epifluorescence microscope (Axioskop 2, Carl Zeiss AG, Germany). The CCD

camera was controlled by using Image-Pro Plus software (version 4.5.1, Media Cybernetics, Yorktown, VA, USA), and the final image adjustments were done with Adobe Photoshop CS2 (version 9.0.2, Adobe Systems Incorporated, San Jose, CA, USA).

4.4 Results

4.4.1 Genome size and chromosome numbers

Genome size estimation in section *Barbata* using flow cytometry (FCM) proved to be technically challenging due to a high proportion of partial endoreduplicated nuclei (data not shown), a phenomenon that has also been noted by Trávníček et al. (2015). This made identification of the correct 2C peak in the FCM histograms difficult. In addition, the tessellated leaves from section *Barbata* and subgenus *Parvisepalum* were found to contain compounds that caused relatively high CV values and low nuclei counts. Pigments and secondary metabolites are problematic as they are known to cause erroneous genome size estimates (reviewed in Greilhuber et al. 2007) as they inhibit the binding of propidium iodide to the DNA. Preliminary tests using various nuclei isolation buffers and plant tissues i.e. pigment-free root tissue, floral bracts and leaves (results not shown) (1) successfully identified the Ebihara (Ebihara et al. 2005) buffer as being the most suitable buffer for genome size estimation and (2) confirmed that the interfering compounds did not affect the accuracy of leaf-based estimates.

In this study, new FCM estimates were obtained for 25 *Paphiopedilum* taxa, 13 of which were from section *Barbata*, the subgroup which displays the greatest chromosome number variation (2*n*=28-42). A summary of these new FCM estimates and published chromosome counts is presented in Table 4.3.

Genomes sizes across subgenus *Parvisepalum*, subgenus *Brachypetalum*, section *Cochlopetalum*, section *Coryopedilum*, section *Pardalopetalum* and section *Paphiopedilum* were not very variable, ranging from 2C=45.7 pg (*P. armeniacum*, 2*n*=26) to 2C=56.2 pg (*P. fairrieanum*, 2*n*=26). In contrast, within section *Barbata*, 2C-values were relatively higher and more varied, with a mean 2C-value of 68.0 pg and a large range from 2C=56.8 pg (*P. bullenianum*, 2*n*=40, 42) to 2C=79.0 pg (*P. sangii*, 2*n*=28), the latter being amongst the largest genome sizes reported for subfamily Cypripedioideae. In addition to *P. sangii*, large (2C>70 pg) and modestly large (2C>65 pg) genome sizes were also found in *P. wardii* (2C=73.5 pg, 2*n*=41), *P. violascens* (2C=70.5 pg, 2*n*=38), *P. purpuratum* (2C=68.3 pg, 2*n*=40) and

P. urbanianum (2C=67.2 pg, 2*n*=40). No clear association between chromosome number and genome size was detected in this study (Fig. 4.1). In addition, a comparison of genome sizes for recently diverged taxa such as the *P. appletonianum-bullenianum* alliance of section *Barbata*, for which there is chromosome number variation (2*n*=38, 40, 42), found only minor differences (0.3 pg) in the genome sizes (Table 4.3).

Generally there was good agreement between the FCM estimates obtained in this work and Feulgen microdensitometry estimates reported in the literature (see Table 4.3). However, for two taxa, *P. barbatum* and *P. callosum*, the FCM and Feulgen values differed substantially by 7.9 and 12.1 pg, respectively. In taxa for which multiple individuals were available, FCM detected only low intra-taxon genome size variation (2C std dev<2.5 pg). Genome sizes between closely related taxa, such as in the *P. appletonianum-bullenianum* (2C=55.7-56.5, 2*n*=38, 40, 42) and *P. barbatum-callosum-lawrenceanum* (2C=59.6-60.3 pg, 2*n*=32, 36, 38) alliances, were generally very similar. However, a surprising difference of >16 pg was observed between the 2C-values of the closely related *P. hookerae* (2C=62.7 pg) and *P. sangii* (2C=79.0 pg) despite these taxa having the same chromosome number (2*n*=28).

4.4.2 Occurrence of repetitive DNA

Illumina HiSeq sequences, corresponding to c. 1.7 to 15.3% of the genome (depending on taxon), for seven *Paphiopedilum* taxa representing the major subgenera/sections of *Paphiopedilum*, and *Phragmipedium longifolium* as the outgroup, were clustered using the RE pipeline. The proportion of repeat types in the various taxa are summarised in Table 4.4. Overall, total repetitive elements accounted for between 61.1 and 71.5% of the genomes in *Paphiopedilum*, the majority of which could be classified into repeat type by RE. Nevertheless, between 9.9 and 12.0% of the genome sequences were identified as repetitive but could not be classified.

Long terminal repeat (LTR) retroelements are the most abundant class of elements in all taxa, accounting for between 41.0 and 55.4% of the genomes. Of the various LTR elements, *Ty3/gypsy* elements, especially of the Ogre/Tat subclass, are by far the most abundant type, comprising up to 51.5% of the genomes studied. Given that the genome sizes vary across the dataset from 2C=13.7 pg in the outgroup (*Phragmipedium longifolium*) to 2C=56.9 pg (*Paphiopedilum appletonianum*), a range of 43.2 pg, the broadly similar proportions of repeats can translate to considerable copy number variations in repeats. In terms of total amount of DNA, the Ogre/Tat class of LTR-elements were the most variable, ranging from 4 pg in *Phragmipedium longifolium* to 23 pg in *Paphiopedilum appletonianum*.

Another major group of LTR-elements found in many plants, the *Ty1/copia* elements, were seen to make up between 1.4 and 5.4% of the *Paphiopedilum* genomes investigated. Of these, Maximus/SIRE and Ivana elements were more abundant in *Paphiopedilum*, whereas Tork elements were the most abundant (7.2%) *Ty1/copia* subclass in *Phragmipedium longifolium*.

Satellite DNA was found to be a significant (>1%) genomic component in all *Paphiopedilum* species analysed with the exception of *P. concolor* (subgenus *Brachypetalum*) where only 0.13% of the genome was estimated to be satellite DNA. Four distinct major types of satellites were identified by RE with varying abundances between taxa. The satellite *SatA* appeared to be specific to *P. armeniacum* (subgenus *Parvisepalum*), where it accounted for 14.4% of the genome. Another satellite, *SatB was* found across all *Paphiopedilum* subgroups ranging from 0.1% in *P. armeniacum* and *P. concolor* (subgenus *Parvisepalum* and section *Brachypetalum* respectively) to 7.5% in *P. lowii* (section *Pardalopetalum*). The satellite *SatG* was found in the genus *Phragmipedium* and most of genus *Paphiopedilum* with the exception of subgenus *Parvisepalum* and subgenus *Brachypetalum* lineages where it appeared to have been lost. *SatJ* was exclusively found in section *Paphiopedilum* and section *Barbata*, in which it comprised 2.5 and 1.8% of the *P. villosum* and *P. appletonianum* genomes, respectively.

As satellites have previously been shown to be useful as chromosome-specific markers (e.g. Koukalova et al. 2010), we further investigated *SatA* by developing a fluorescence *in situ* hybridisation (FISH) probe designed from *SatA* sequence contigs that had been assembled by RE (see section 4.3.5). A FISH experiment conducted by Dr. Yung-I Lee confirmed that this satellite was indeed specific to subgenus *Parvisepalum* (Fig. 4.3 and 4.4). In *P. armeniacum*, the marker gave multiple localised signals across the karyotype.

Other significant repeat elements include DNA transposons and LINE elements which respectively comprised 1.3 to 2.4% and 0.1 to 0.6% of *Paphiopedilum* genomes. There were two additional, surprising results: (1) rDNA, which must be present in all species, was not detected in three of the species analysed, presumably reflecting the sensitivity of the method, and (2) there was a high proportion (28.5-38.9%) of the genome that was single- or low-copy repetitive DNA.

4.4.4 Using repeat abundances to build a phylogenetic tree

The most parsimonious tree, constructed using the relative abundances of 1000 RE clusters as continuously varying characters, from the comparative analysis of sequences comprising

approximately 0.1% of the genome for 14 representative *Paphiopedilum* and a *Phragmipedium* outgroup species, is presented in Fig. 4.5 alongside plastid and ITS trees derived from the same samples for comparison.

The established groupings of section *Barbata*, section *Paphiopedilum*, subgenus *Brachypetalum* and subgenus *Parvisepalum*, which are supported on the plastid and ITS trees, were also recovered on the RE tree with good support (BP≥90) (Fig. 4.5). However, the multi-flowered groups, comprised of sections *Coryopedilum*, *Pardalopetalum* and *Cochlopetalum*, for which there were gene tree incongruences and paraphyly (see Chapter 2, section 2.5.1), were not resolved as monophyletic on the RE tree. This is in contrast to the plastid and ITS trees, both of which resolved a clade containing all three sections, implying a close relationship between the multiflowering groups, although the precise inter-sectional relationships differed. A supernetwork of 10% of the RE bootstrap trees (Fig. 4.6) visualises the conflicting placements in these groups.

The unusual taxon *P. fairrieanum*, for which there is conflicting phylogenetic placement (see Chapter 2, section 2.5.1), was resolved with good support (BP=93) within section *Paphiopedilum* on the RE tree. This agrees with the ITS tree where *P. fairrieanum* was placed with section *Paphiopedilum*, albeit with only weak support (BP=57). However, a different relationship was suggested by the plastid tree, which resolved *P. fairrieanum* as sister to a clade containing section *Barbata*, section *Paphiopedilum*, section *Coryopedilum*, section *Cochlopetalum* and section *Pardalopetalum* with modest support (BP=85).

4.5 Discussion

4.5.1 Genome size and chromosome number evolution

The study presented here has extended current knowledge of genome size evolution and provided new FCM-based genome size estimates for *Paphiopedilum*. The new data include the largest *Paphiopedilum* genome so far reported for the genus in *P. sangii* (2*n*=28, 2C=79 pg). It is noted that the genome size of *P. sangii* is larger than that of the previous record holder in Cypripedioideae, which was for *Cypripedium henryi* (2C=77.7 pg; Bennett et al. 2000). Indeed, *P. sangii* has the second largest genome so far reported for any orchid (the largest being *Pogonia ophioglossoides* in the subfamily Vanilloideae; (Leitch et al. 2009). Even within *Paphiopedilum*, the genome of *P. sangii* is approximately 30% larger than its closest relative (*P. hookerae*, 2*n*=28, 2C=62.7 pg), thus suggesting that a large expansion in genome size of c. 16,000 Mb of DNA may have taken place within an evolutionarily short period of time (probably less than 3 million years, based on the dating analysis given in Chapter 3, see section

3.5.1 and Fig. 3.4). Further work characterising the genome of *P. sangii* is needed to shed light on the nature of this genome size increase e.g. what types of DNA sequences are involved and where in the karyotype the additional DNA is being added. Overall, the new data have extended the range of genome sizes encountered in *Paphiopedilum* to 2.4-fold from 2C=33 pg in *P. exul* (Bennett et al. 2000) to 2C=79 pg in *P. sangii*.

Earlier assessments by other workers (Cox et al. 1998; Chochai et al. 2012) suggested an ancestral karyotype of 26 metacentric chromosomes (2*n*=2*x*=26) in *Paphiopedilum* with a general trend towards increasing chromosome numbers independent of polyploidy accompanying the radiation of the genus. Although there is a similar trend toward increasing chromosome number within section *Barbata* (2*n*=28-42), analysis of the additional data reported here failed to uncover any clear trend between chromosome number increase and genome size either in section *Barbata* (Fig. 4.1) or in the genus *Paphiopedilum* as a whole (Chochai et al. 2012). Such results are similar to those previously reported in *Carex scoparia* (Cyperaceae) (Chung et al. 2011) but are in contrast with those of de Azkue and Martínez (1988) who noted that an increase in chromosome number by centric fission in the genus *Oxalis* was accompanied by an increase in DNA amount.

The observation that many taxa of section *Barbata* with high chromosome numbers occur on islands led Cox et al. (1998) to suggest that colonisation of islands may have triggered chromosome number increases. However, the phylogenetic data from plastid and two low-copy nuclear genes presented in Chapter 2 suggest that increases in chromosome number appear to have arisen multiple times within section *Barbata* among island as well as in continental taxa e.g. *P. bullenianum* var. *celebense* (2*n*=42) from Wallacea, *P. urbanianum* (2*n*=40) from the Philippines, *P. purpuratum* (2*n*=40) and *P. wardii* (2*n*=41, 44) from northern Indochina.

Available data suggest that the chromosome number of 2*n*=26 is conserved in subgenus *Parvisepalum*, subgenus *Brachypetalum*, section *Coryopedilum*, section *Pardalopetalum*, section *Paphiopedilum* and the recently described subgenus *Megastaminodium* (see Table 1.5). Higher chromosome numbers and total number of chromosome arms, i.e. nombre fundamental (*n.f.,* Duncan and Macleod 1949; Matthey 1949) occur in section *Barbata* (2*n*=2*x*=28-42, *n.f.*=52, 53, 56) and section *Cochlopetalum* (2*n*=2*x*=30-37, *n.f.*=48, 50). Based on the general conservation of the *n.f.* and a molecular ITS-based phylogenetic tree (Cox et al. 1998), previous workers posited that the range of chromosome numbers in *Paphiopedilum* was predominantly caused by centric fissions and fusions. However, more recent work suggests that this chromosome evolution model is too simple. For

example, a fluorescence *in situ* hybridisation (FISH) study by Lan and Albert (2011) using 25S and 5S rDNA probes revealed complex rDNA signals with considerable diversity in the number and location of the different rDNA sites even between closely related species. This led the authors to conclude that in addition to chromosome fissions, more complex rearrangements, including inversions, may also have taken place during the evolution of *Paphiopedilum*. Nevertheless, the limited information from the two rDNA FISH markers and the complex dynamics of rDNA evolution did not allow the more precise identification of rearrangement types to specific taxa. More work with additional FISH markers and other genetic tools, e.g. large scale sequencing and mapping, is needed to properly characterise the genomes of *Paphiopedilum* and hence gain insight into the full complement of genomic processes operating.

Yet another layer of complexity in the dynamics of genomic changes was added by recent molecular phylogenetic studies using low-copy nuclear genes and plastid markers (see Chapter 2; Górniak et al. 2014; Guo et al. 2015). These studies implicate historical homoploid hybridisation or reticulation in the evolution of *Paphiopedilum*, specifically in section *Barbata*, section *Cochlopetalum*, section *Paphiopedilum* and possibly subgenus *Megastaminodium*. Certainly there is growing evidence (reviewed in Ågren and Wright 2011) that homoploid hybridisation between divergent genomes can trigger genomic phenomena such as chromosome gains or losses, modified chromosomal pairing and recombination and transposable element activation, and it seems plausible that these processes may underpin the diversity of genome sizes and chromosome numbers, particularly in section *Barbata*.

In Chapter 2, reticulation was uncovered between taxa with differing chromosome numbers within section *Barbata*. For instance, the conflicting gene trees (Fig. 2.5A-B, Fig. 2.6A-B and 2.7A-B) suggest reticulation between *P. barbatum* (2*n*=38, *n.f.*=52) and *P. callosum* (2*n*=32, *n.f.*=52) in the ancestry of *P. lawrenceanum* (2*n*=36, *n.f.*=52). Introgressive hybridisation is also suggested to have occurred between *P. callosum* and *P. appletonianum* (2*n*=38, *n.f.*=52), between *P. barbatum* and *P. bullenianum* (2*n*=40, *n.f.*=52) and between *P. bullenianum* and *P. javanicum* (2*n*=38, *n.f.*=52). Interestingly, it is noted that all of the cases of reticulation that have been uncovered thus far have been between taxa with similar genome sizes (<5 pg difference). Collectively, this suggests that (1) genome sizes may be conserved between taxa that hybridise frequently, similar to the situation reported for Fagaceae (Chen et al. 2014) and that (2) large differences in genome sizes may be barriers to hybridisation similar to the situation observed in *Phalaenopsis* (Orchidaceae) (Chen et al. 2013). The taxon *P. wardii* (2*n*=41, *n.f.*=53), which sometimes carries an aneuploid chromosome, is suspected to have a reticulate origin (Chapter 2; Karasawa 1986).

Reticulation signals between sympatric taxa were detected throughout section *Barbata* (Chapter 2, section 2.5.2). However, they were absent in *P. hookerae* and *P. sangii*, despite the fact that they exist in sympatry with other taxa of section *Barbata*. These two taxa share a common unique karyotype (2*n*=28, *n.f.*=56) within section *Barbata* and their sister relationship is strongly supported across all gene trees (Fig. 2.5A-B, Fig. 2.6A-B, Fig. 2.7A-B). We hypothesise that the karyotype structure in these two species may potentially contain rearrangements, such as large inversions, that prevent hybridisation with other karyotypes. Further work will need to be done to characterise the genomic constituents of these two taxa.

4.5.2 Characterisation of repetitive elements

While some cases of genome size differences in *Paphiopedilum* can reasonably be explained as a gain or loss of DNA as a by-product of chromosome fissions (Werner et al. 1992, Tsujimoto et al. 1997, Putnam et al. 2004), translocations (Nishikawa et al. 1979; Pijnacker and Schotsman 1988) and hybridisation (Baack et al. 2005; Marques et al. 2012), there is no evidence that any one of these processes can account for the overall trend of genome size expansion observed in *Paphiopedilum*. In this study, we used RE to characterise the genomes from six *Paphiopedilum* representatives to understand the basic changes taking place in *Paphiopedilum* genomes better.

RE analysis revealed that all *Paphiopedilum* subgroups contained over 60% of repetitive DNA (Table 4.4). The repetitive DNA fraction was found to be composed of many types of transposable elements across all *Paphiopedilum* subgroups. However, no association between the expansion of genome size and the proliferation of any one type of class of repetitive element was observed. Rather, genome size expansion appeared to be caused by an overall expansion of multiple elements, similar to the situation observed in *Musa* (Musaceae) (Novák et al. 2014) and *Fritillaria* (Liliaceae) (Kelly et al. 2015).

In addition to transposable elements, tandem repeats are another commonly found type of repetitive element in plant genomes. These sequences represent DNA of a few nucleotides to many hundreds of nucleotides that typically occur in a tandem head to tail organization. When the repeat unit is small (i.e. <120 bp), these are referred to as minisatellite repeats, and when longer (typically around 120-180 bp), they are frequently referred to as satellite repeats. Several types of satellite DNA were detected across all *Paphiopedilum* subgroups, but they were especially abundant in section *Barbata*, section *Pardalopetalum* and subgenus *Parvisepalum*. FISH results using a probe derived from *SatA*, which was specific to subgenus *Parvisepalum*, gave localised and heterogeneous signals on subtelomeric and

centromeric chromosome regions (Fig. 4.3) which are consistent with findings from other studies (e.g. Koukalova et al. 2010). This distribution pattern of the satellite revealed by FISH is valuable as it potentially allows it to be used as a chromosome-specific marker for tracking chromosomal rearrangements in *Paphiopedilum*.

A surprisingly large (approximately 30 to 40%) proportion of *Paphiopedilum* genomes was found to consist of single- and low-copy DNA, representing between 13 and 20 Gbp of DNA. In comparison with the *Arabidopsis thaliana* 2C genome size, which is about 0.3 Gb (Bennett et al. 2003), the estimated amount of single- and low-copy DNA in *Paphiopedilum* amounts to between 43 and 67 whole genomes of *A. thaliana* that is single- or low-copy. This finding was a surprise as one might have predicted that the extra DNA found in *Paphiopedilum* relative to *A. thaliana* would be entirely due to repetitive DNA. Nevertheless, studies on *Arabidopsis* (Maumus and Quesneville 2014) and *Fritillaria* (Kelly et al. 2015) suggest that a large proportion of single- and low-copy copy DNA may be derived from ancient degraded repetitive elements. If similar mechanisms are operating in *Paphiopedilum*, it is possible that the single and low-copy repeats are essentially the remains of very old repetitive elements that have accumulated sufficient mutations over time so that they are no longer recognised as repetitive. This in turn suggests that genome expansion observed in *Paphiopedilum* is due to a low frequency of repetitive DNA removal rather than high rate of repetitive element amplification.

A second surprising result was the apparent absence of rDNA sequences in many species and the low proportion in all species, the maximum rDNA volume being 0.02 Mb DNA (2C) which, if we assume the rDNA unit is 10 kb long, accounts for only three rDNA copies per diploid genome. Clearly such low numbers cannot be the case, since rDNA is essential and expected in hundreds of copies, based on studies of other plant genomes and results from FISH. The data presented here suggest that: (1) rDNA is unrepresented in Illumina sequences (i.e. it is a technical artefact); (2) RE fails to recognise efficiently these repeats in clusters (although this is not reported for other species); or (3) the copy number of rDNA is low in *Paphiopedilum*, although this is not suggested by the FISH data of Lan and Albert (2011). Given that we analysed volumes of Illumina data that were equivalent to only up to 16% of the genome (Table 4.2), we can only expect to find a repeat in copy numbers greater than about 20-100 copies. However, previous work (Lan and Albert 2011) has suggested that *Paphiopedilum* has numerous (although variable numbers of) rDNA loci, and strong signals, both observations would suggest that several hundred rDNA copies should be expected and hence the failure to find rDNA sequences in the Illumina data may indeed be a sequencing artefact.

4.5.3 Evaluation of the RE cluster-based phylogenetic analysis

I investigated the potential utility of the recently described RE cluster-based technique developed by Dodsworth et al. (2015) for phylogenetic analysis of *Paphiopedilum*. RE-based comparative analysis represents an attractive method for evaluating phylogenetic relationships as it utilises the often-ignored by-products of other sequencing work e.g. those focused on plastid and mitochondria genome assembly. Working with organisms such as *Paphiopedilum* presented a challenge for RE comparative analysis due to the large genome size of the group which is computationally demanding and limits the number of samples that can be effectively clustered at one time. In this study we sampled approximately 0.2% of the genome proportion (GP) for 14 *Paphiopedilum* taxa and a *Phragmipedium* outgroup which is the upper limit currently allowed by the RE pipeline while still falling within the recommended GP range (Dodsworth et al. 2015). Nevertheless, it should be noted that low-abundance clusters may give misleading phylogenetic signals due to the sensitivity of RE to GP i.e. the lower the actual abundance of the cluster in the genome, the higher the inherent error in its estimated abundance will be due to chance. In part, this effect was minimised by the implementation of the character scoring method of Dodsworth et al. (2015) which gives greater weight to the highly abundant clusters.

Comparisons of the tree produced from RE with plastid- and nuclear ITS-based (Fig. 4.5) trees suggest that the RE cluster-based phylogenetic analysis is broadly congruent with trees produced from molecular analysis of plastid and nuclear ITS loci. Clades which were unequivocally supported by both plastid and nuclear loci were also recovered with strong support on the RE tree. However, clades for which there exist gene tree conflicts or paraphyly on plastid and nuclear loci were resolved in agreement with one of the trees, as paraphyletic or weakly supported.

Although preliminary in nature, this study suggests that RE-based phylogenetic analysis may present yet another method for assessing the evolutionary relationships in *Paphiopedilum*. Further comparative work with differing combinations of samples at different phylogenetic levels e.g. below section-level relationships etc., comparisons with phylogenetic trees from more molecular loci, using different methods e.g. different character scoring methods, will be needed to provide a more robust assessment of the efficacy of this method.

4.6 Conclusions

In this study we have added valuable new FCM genome size data and repetitive DNA genomic data for *Paphiopedilum* section *Barbata*. Although conclusive evidence of precise chromosomal rearrangements are still lacking, the findings from this study interpreted along with data from other recently published studies (Lan and Albert 2011; Guo et al. 2015) have nonetheless enabled us to reassess and update existing theories on genome evolution in *Paphiopedilum,* especially within section *Barbata*, the subgroup which shows the greatest range of chromosome numbers and genome sizes.

I show that genome evolution in *Paphiopedilum* is characterised by multiple independent chromosome number increases and genome expansions. We hypothesise that these changes are driven, not only by simple chromosome fissions and translocations as was originally proposed (Cox et al. 1998), but also by hybridisation between *Paphiopedilum* taxa with divergent genome lineages, and possibly including inversions in specific lineages as proposed here for *P. hookerae* and *P. sangii*. Remarkably, we found that genome size expansion in *Paphiopedilum* was not due to a proliferation of any one specific family of repetitive elements. Rather, it appeared to be caused by an overall increase in all types of repetitive elements, suggesting that the changes may be due to alterations in a general regulatory mechanism that inhibits repetitive DNA removal such as epigenetic pathways which play an important role in regulating the activity and elimination of repetitive DNA (Bennetzen and Wang 2014; Matzke and Mosher 2014). In addition, we were able to use the RE data to develop a FISH marker from a satellite repeat element identified in *P. armeniacum*. FISH results suggest that this marker shows potential as a chromosome-specific marker for studying rearrangements in subgenus *Parvisepalum*.

Comparative analysis of the repetitive DNA of 15 HiSeq samples representing the major *Paphiopedilum* subgroups and a *Phragmipedium* outgroup allowed us to evaluate a newly developed phylogenetic method which uses repetitive element graph clusters as continuously varying characters. Our findings suggest that the method produces logical topologies that are largely congruent with those obtained from plastid- and nuclear-based molecular markers in *Paphiopedilum*. Although preliminary in nature, the findings from this study show that the RE-based method can provide useful phylogenetic information for assessing relationships in *Paphiopedilum*.

Tables

Table 4.1 List of plant materials used for genome size estimation

Taxon	Source	Verification (P/V)	Sample reference
Paphiopedilum			
section *Barbata*			
P. appletonianum A	QM	Pending	QM3020
P. appletonianum B	QM	Pending	QM3023
P. barbatum	K	P	K20101538
P. bullenianum A	K	V & P	K1992654
P. bullenianum B	QM	Pending	QM3003
P. bullenianum C	QM	Pending	QM3026
P. purpuratum	HBL	V & P	HBL 20110252
P. schoseri	QM	V & P	QM3007
subgenus *Brachypetalum*			
P. concolor	HBL	V & P	HBL 20110263
section *Cochlopetalum*			
P. primulinum	K	P	K19811628
section *Coryopedilum*			
P. glanduliferum	HBL	V & P	HBL 1071
P. rothschildianum	K	V	K19844033/36806
section *Paphiopedilum*			
P. druryi†	HBL	P	Not available
P. fairrieanum	HBL	V & P	HBL 20110265
P. henryanum	HBL	V & P	HBL 20110255
P. villosum A	K	Pending	K20052739
P. villosum B	K	Pending	K20101520
section Pardalopetalum			
P. lowii	HBL	V	HBL 30629
subgenus *Parvisepalum*			
P. armeniacum	MBG	V & P	X-0779
Phragmipedium			
Phragmipedium longifolium	HBL	V & P	HBL 20110235

Abbreviations:

Source

HBL - Hortus botanicus of Leiden University, Netherlands
K – Royal Botanic Gardens, Kew, UK
MBG – Munich Botanic Gardens, Germany
QM – Queen Mary University of London, UK

Verification

V – Voucher
P – Photograph

Table 4.2 List of plant materials used for *HiSeq* sequencing and *RepeatExplorer* (RE) cluster analyses

Taxon	Source	Genome proportion used for RE cluster characterisation (%)	Genome proportion used for RE comparative analysis (%)	Verification (P/V)	Sample reference
Paphiopedilum					
section *Barbata*					
P. appletonianum	TNM	1.68	0.22	V	Yung-I Lee 201209
P. barbatum	HBL	NA	0.22	V	HBL 200070142
P. purpuratum	HBL	NA	0.22	V & P	HBL 20110252
subgenus *Brachypetalum*					
P. concolor	TNM	4.77	0.22	V	Yung-I Lee 201201
section *Cochlopetalum*					
P. primulinum	K	2.27	0.22	P	K19811628
section *Coryopedilum*					
P. glanduliferum A	HBL	NA	0.22	V & P	HBL 1071
P. glanduliferum B	HBL	NA	0.22	V	KAS 21 1071
P. rothschildianum	TNM	6.27	0.22	V	Yung-I Lee 201107
section *Paphiopedilum*					
P. druryi	HBL	NA	0.22	P	HBL 2013 0012
P. fairrieanum	HBL	NA	0.22	V & P	HBL 20110265
P. henryanum	HBL	NA	0.22	V & P	HBL 20110255
P. villosum	TNM	3.53	0.22	V	Yung-I Lee 201012

(cont.)

Table 4.2 (cont.)

Taxon	Source	Genome proportion used for RE cluster characterisation (%)	Genome proportion used for RE comparative analysis (%)	Verification (P/V)	Sample reference
***Paphiopedilum* (cont.)**					
section *Pardalopetalum*					
P. lowii	HBL	5.22	0.22	V	HBL 30629
subgenus *Parvisepalum*					
P. armeniacum	TNM	3.48	0.26	V	Yung-I Lee 201101
Phragmipedium					
Phragmipedium longifolium	HBL	15.30	0.22	V & P	HBL 20110235

Abbreviations:

Source

HBL – Hortus botanicus of Leiden University, Netherlands

K – Royal Botanic Gardens, Kew, UK

TNM – National Museum of Natural Science, Taichung, Taiwan

Verification

V – Voucher

P – Photograph

Table 4.3 Flow cytometry genome size estimates and published chromosome numbers for *Paphiopedilum*

Taxon	Chromosome number (2*n*)	Chromosome number reference	No. of individuals	FCM 2C (mean±std dev)	Feulgen 2C***	Feulgen 2C reference
Paphiopedilum						
section *Barbata*						
P. appletonianum	38	Karasawa 1979	2	56.48±0.56	54.05	Bennett and Leitch 1997
P. barbatum	38	Karasawa 1979	1	59.59	67.50	Bennett and Leitch 1997
P. bullenianum	40, 42*	Karasawa 1979*, 1986; Karasawa and Aoyama 1980	3	55.7±0.76	51.70	Bennett et al. 2000
P. callosum	32	Karasawa 1979	3	60.31±1.77	48.15	Bennett and Leitch 1997
P. hookerae	28	Karasawa 1979	1	62.72		
P. lawrenceanum	36	Karasawa 1979	2	60.27±2.02		
P. purpuratum	40	Karasawa 1979	1	68.32		
P. sangii	28	Karasawa et al. 1997	1	78.95		
P. schoseri	35	Karasawa et al. 1997	1	57.35		
P. tonsum	32	Karasawa 1979; Cox et al. 1998	1	57.47		
P. urbanianum	40	Karasawa 1982	1	67.20		
P. violascens	38	Karasawa 1979	1	70.47		
P. wardii	41	Karasawa 1986	2	73.50±0.17		
subgenus *Brachypetalum*						
P. concolor	26	Karasawa 1979, 1986	1	46.19		
section *Cochlopetalum*						
P. primulinum	32	Karasawa 1979	1	47.12		
section *Coryopedilum*						
P. glanduliferum	26	Karasawa 1979; Cox et al. 1998	1	54.76		
P. rothschildianum	26	Karasawa 1979	1	48.02		

(cont.)

Table 4.3 (cont.)

Taxon	Chromosome number (2*n*)	Chromosome number reference	No. of individuals	FCM 2C (mean±std dev)	Feulgen 2C***	Feulgen 2C reference
***Paphiopedilum* (cont.)**						
section *Paphiopedilum*						
P. druryi	30	Karasawa 1979	1	50.08	53.50	Bennett and Leitch 1997
P. fairrieanum	26	Karasawa and Saito 1982	1	56.17		
P. henryanum	26	Karasawa and Aoyama 1988	1	49.48		
P. villosum	26	Karasawa 1979	1	49.53		
section *Pardalopetalum*						
P. lowii	26	Karasawa 1979	1	53.96	49.05	Bennett et al. 2000
subgenus *Parvisepalum*						
P. armeniacum	26	Karasawa 1982	1	45.74		
P. emersonii	26	Karasawa 1986	1	48.29		
Phragmipedium						
Phragmipedium longifolium	20+3f*, 21**	Karasawa 1980*; Cox et al. 1998**	1	13.68	12.20	Bennett and Leitch 1997

***Only vouchered Feulgen samples were included in this table

Table 4.4 Genome proportions of repetitive elements in *Paphiopedilum*

	Outgroup		***Parvisepalum***		***Brachypetalum***		***Cochlopetalum***		***Coryopedilum***		***Pardalopetalum***		***Paphiopedilum***		***Barbata***	
	Phragmipedium longifolium ***2C=13.7 pg***		*P. armeniacum* ***2C=45.7 pg***		*P. concolor* ***2C=46.2 pg***		*P. primulinum* ***2C=47.1 pg***		*P. rothschildianum* ***2C=48.0 pg***		*P. lowii* ***2C=54.0 pg***		*P. villosum* ***2C=49.5 pg***		*P. appletonianum* ***2C=56.5 pg***	
	%	**pg**	**%**	**pg**	**%**	**pg**	**%**	**pg**	**%**	**pg**	**%**	**Pg**	**%**	**pg**	**%**	**pg**
LTR elements																
Ty3/gypsy																
Ogre/Tat	30.04	4.12	31.80	14.53	48.86	22.58	38.03	17.91	44.98	21.59	39.06	21.09	46.32	22.93	40.48	22.87
Chromovirus	7.82	1.07	2.49	1.14	1.47	0.68	0.59	0.28	1.44	0.69	0.81	0.43	0.89	0.44	0.66	0.37
Athila	0.20	0.03	4.41	2.01	1.21	0.56	0.21	0.10	1.19	0.57	0.18	0.10	0.28	0.14	0.14	0.08
Total Ty3/gypsy	**38.06**	**5.21**	**38.70**	**17.69**	**51.54**	**23.81**	**38.83**	**18.29**	**47.60**	**22.85**	**40.04**	**21.62**	**47.49**	**23.51**	**41.27**	**23.32**
Ty1/copia																
Maximus/SIRE	3.05	0.42	2.24	1.02	1.56	0.72	0.97	0.46	1.49	0.72	0.48	0.26	0.99	0.49	0.62	0.35
Ivana	0.00	0.00	3.14	1.44	2.14	0.99	1.14	0.54	1.92	0.92	0.81	0.44	2.74	1.36	1.82	1.03
Tork	7.17	0.98	0.00	0.00	0.16	0.08	0.05	0.03	0.04	0.02	0.08	0.05	0.05	0.02	0.07	0.04
Total Ty1/copia	**10.22**	**1.40**	**5.38**	**2.46**	**3.87**	**1.79**	**2.16**	**1.02**	**3.45**	**1.65**	**1.37**	**0.74**	**3.78**	**1.87**	**2.52**	**1.42**
Satellite																
SatA	0.00	0.00	14.39	6.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SatB	0.00	0.00	0.10	0.05	0.13	0.06	6.18	2.91	2.81	1.35	7.45	4.02	1.07	0.53	4.11	2.32
SatG	0.04	0.01	0.00	0.00	0.00	0.00	0.46	0.22	0.65	0.31	1.63	0.88	0.69	0.34	0.42	0.24
SatJ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.51	1.24	1.80	1.02
Total Satellite	**0.04**	**0.01**	**14.49**	**6.62**	**0.13**	**0.06**	**6.64**	**3.13**	**3.47**	**1.66**	**9.08**	**4.91**	**4.27**	**2.12**	**6.33**	**3.58**
Other repetitive elements																
TRIM	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
LINE	**1.09**	**0.15**	**0.57**	**0.26**	**0.49**	**0.23**	**0.10**	**0.05**	**0.76**	**0.36**	**0.53**	**0.28**	**0.36**	**0.18**	**0.26**	**0.14**

(cont.)

Table 4.4 (cont.)

	Outgroup		***Parvisepalum***		***Brachypetalum***		***Cochlopetalum***		***Coryopedilum***		***Pardalopetalum***		***Paphiopedilum***		***Barbata***	
	Phragmipedium longifolium ***2C=13.7 pg***		*P. armeniacum* ***2C=45.7 pg***		*P. concolor* ***2C=46.2 pg***		*P. primulinum* ***2C=47.1 pg***		*P. rothschildianum* ***2C=48.0 pg***		*P. lowii* ***2C=54.0 pg***		*P. villosum* ***2C=49.5 pg***		*P. appletonianum* ***2C=56.5 pg***	
	%	**pg**	**%**	**pg**	**%**	**pg**	**%**	**pg**	**%**	**pg**	**%**	**pg**	**%**	**pg**	**%**	**pg**
Other repetitive elements (cont.)																
DNA transposon	**0.60**	**0.08**	**2.41**	**1.10**	**1.97**	**0.91**	**1.31**	**0.62**	**2.02**	**0.97**	**1.25**	**0.67**	**1.63**	**0.81**	**1.83**	**1.04**
MITE	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
rDNA	**0.14**	**0.02**	**0.04**	**0.02**	**0.04**	**0.02**	**0.06**	**0.03**	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.01**	**0.00**	**0.00**
SSR	**0.23**	**0.03**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Pararetrovirus	**0.00**	**0.00**	**0.08**	**0.04**	**0.00**	**0.00**	**0.00**	**0.00**	**0.16**	**0.07**	**0.00**	**0.00**	**0.00**	**0.00**	**0.59**	**0.33**
Unclassified repetitive	**17.31**	**2.37**	**9.86**	**4.51**	**10.6**	**4.89**	**11.94**	**5.62**	**10.36**	**4.97**	**9.9**	**5.35**	**10.63**	**5.26**	**11.38**	**6.43**
Low and single copy	**32.26**	**4.42**	**28.46**	**13.00**	**31.37**	**14.49**	**38.94**	**18.34**	**32.14**	**15.43**	**37.80**	**20.41**	**31.82**	**15.75**	**35.83**	**20.24**
% TOTAL REPETITIVE DNA	**67.74**	**9.28**	**71.54**	**32.70**	**68.63**	**31.71**	**61.06**	**28.76**	**61.06**	**28.76**	**62.20**	**33.59**	**68.18**	**33.75**	**64.17**	**36.51**

Figures

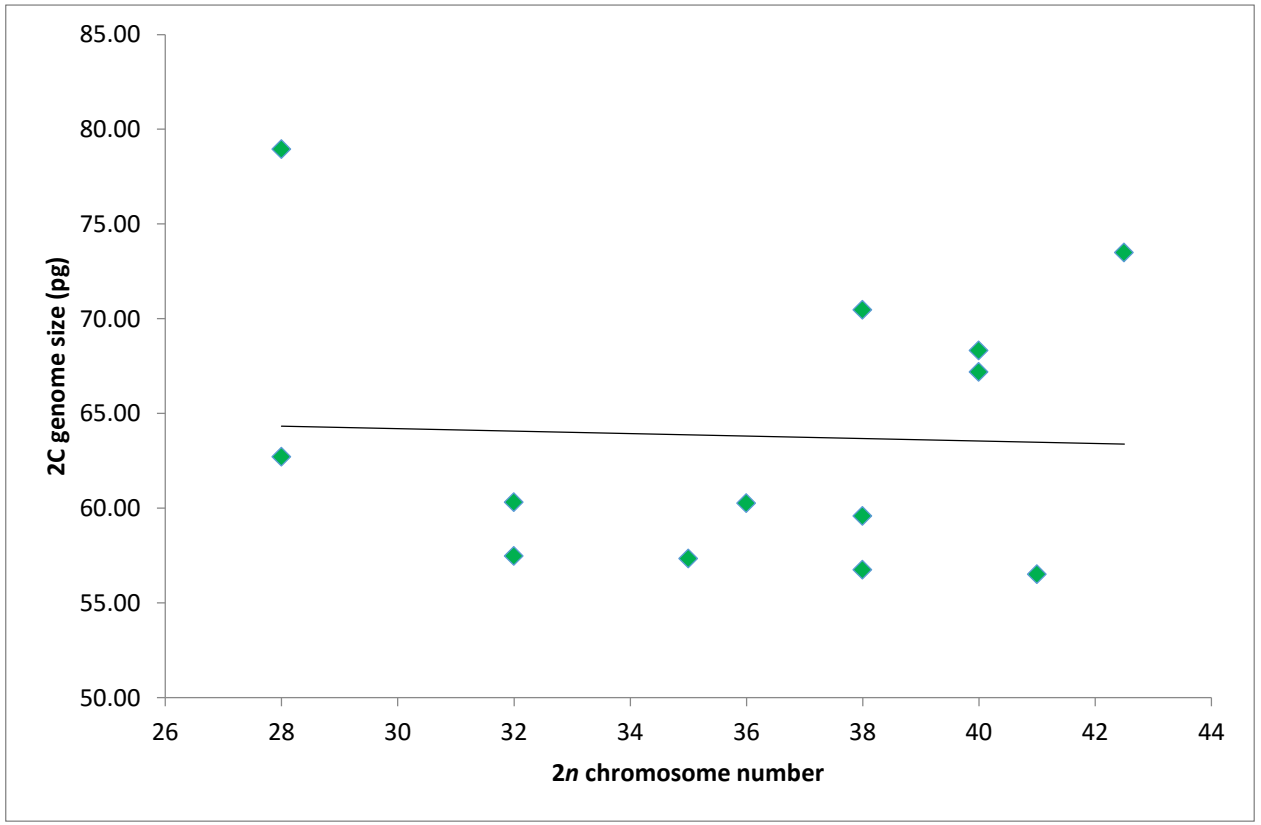

Figure 4.1 Plot of the 2*n* chromosome number against the 2C genome size in section *Barbata*.

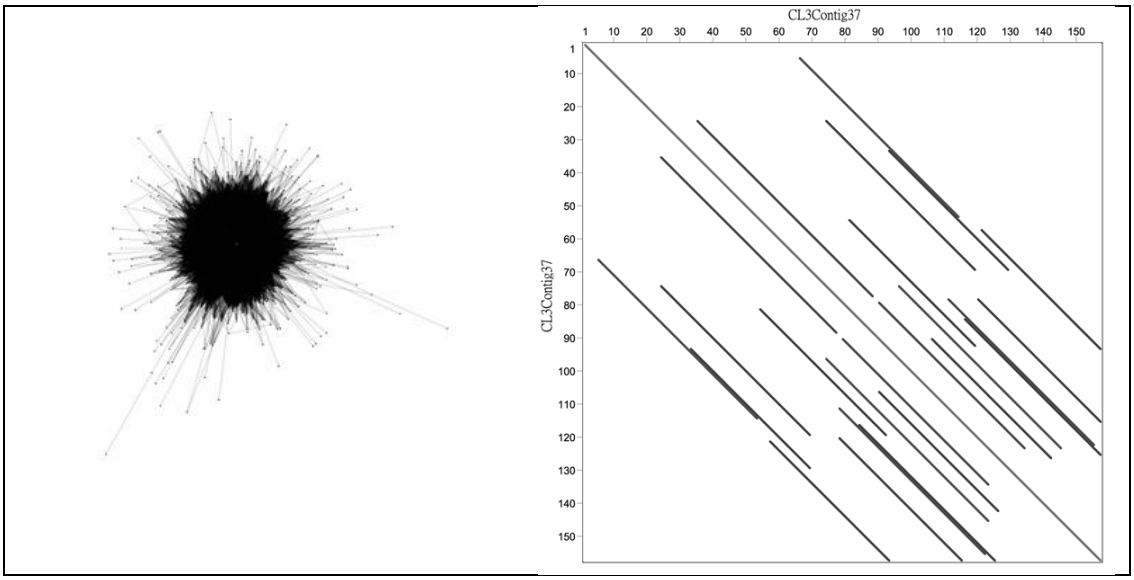

Figure 4.2 The satellite cluster graph of *SatA* from RepeatExplorer with an example of a dot plot of the most abundant *SatA* sequence contig against itself from DOTTER (Sonnhammer and Durbin 1995).

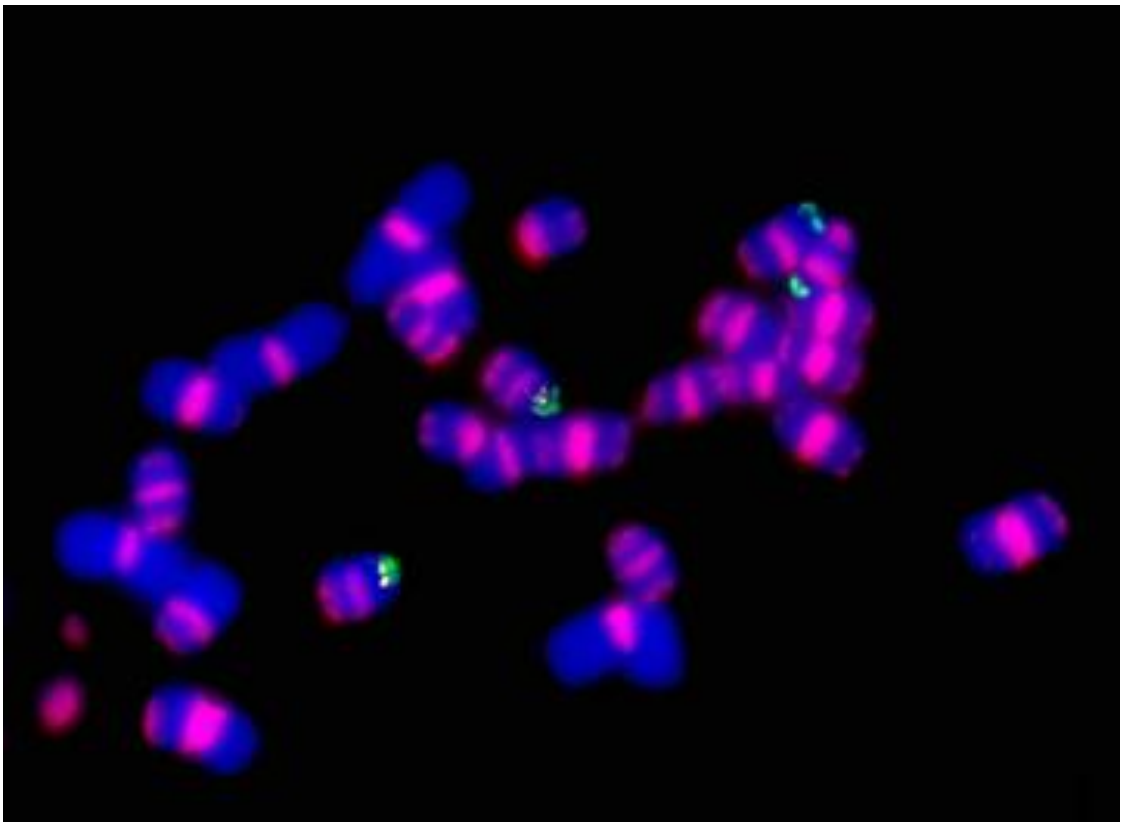

Figure 4.3 Fluorescence *in situ* hybridisation (FISH) of *Paphiopedilum armeniacum* root tip metaphase chromosomes with the *SatA* probe (reproduced from Yung-I Lee, unpublished manuscript). Signals are visualised as *SatA* (pink), 45S rDNA (green) and 5S rDNA (white) against a DAPI counterstain (blue).

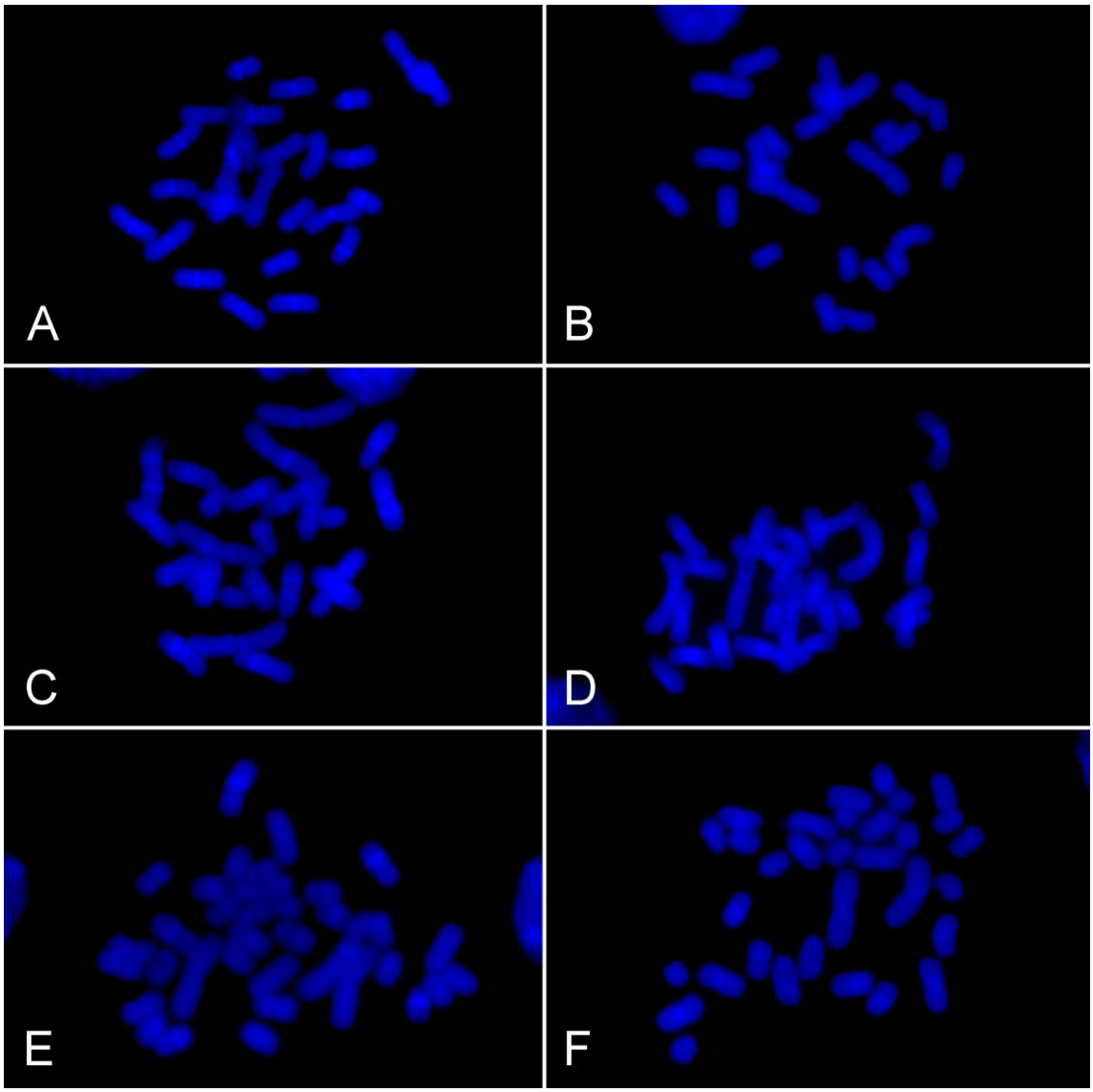

Figure 4.4 Fluorescence *in situ* hybridisation (FISH) of root tip metaphase chromosomes with the *Sat A* probe from Fig. 4.3 in **(A)** *P. concolor* (subgenus *Brachypetalum*), **(B)** *P. villosum* (section *Paphiopedilum*), **(C)** *P. rothschildianum* (section *Coryopedilum*), **(D)** *P. lowii* (section *Pardalopetalum*), **(E)** *P. appletonianum*, and **(F)** *P. primulinum* (section *Cochlopetalum*) (reproduced from Yung-I Lee, unpublished manuscript). The absence of FISH signals confirms that *Sat A* is indeed specific to subgenus *Parvisepalum*.

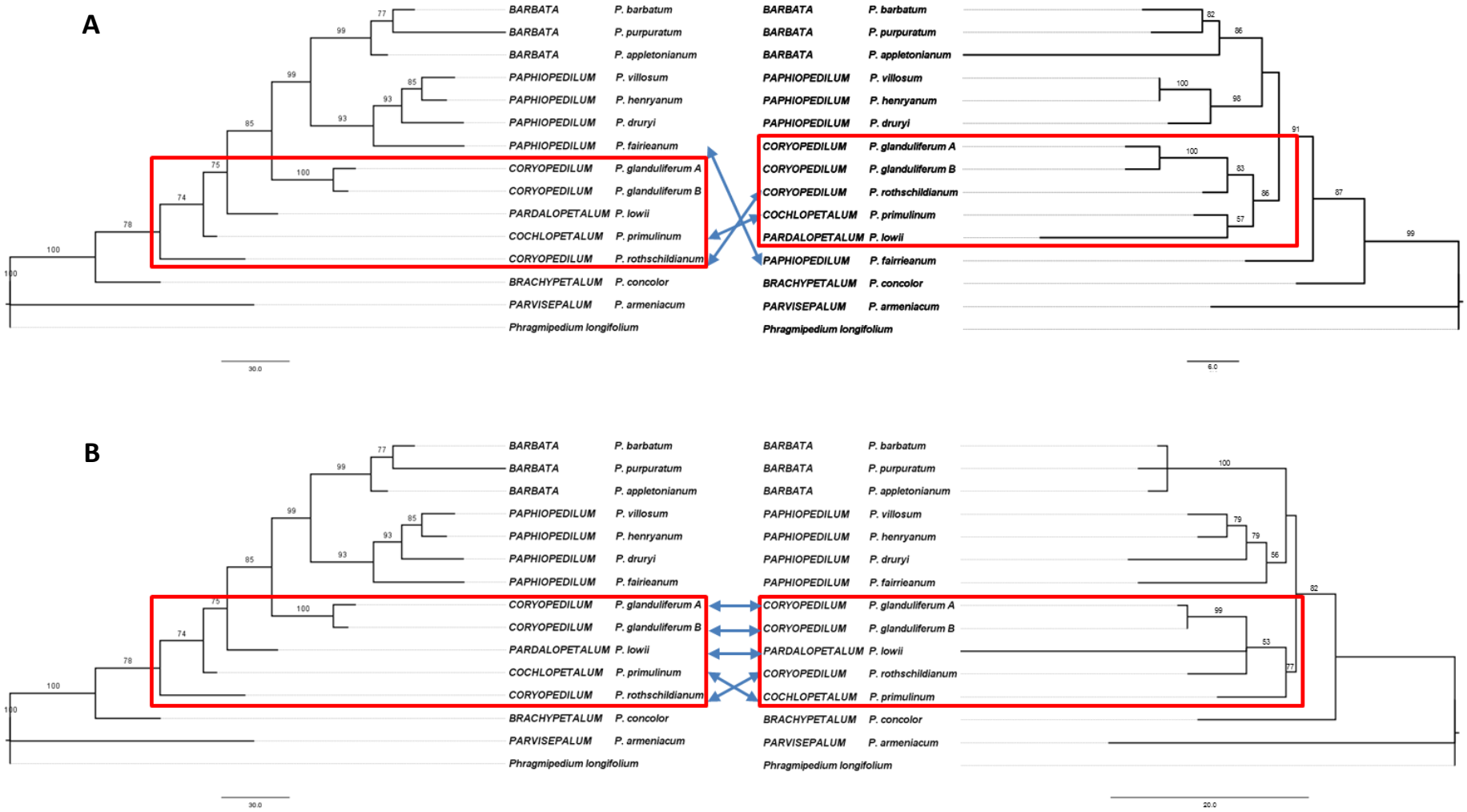

Figure 4.5 Comparison of the most parsimonious RE tree against **(A)** one of three equally parsimonious plastid (*ycf1* & *matK*) trees and **(B)** one of 34 equally parsimonious nuclear ITS trees for 14 *Paphiopedilum* species with *Phragmipedium* as an outgroup. Numbers above branches indicate bootstrap percentages (BP).

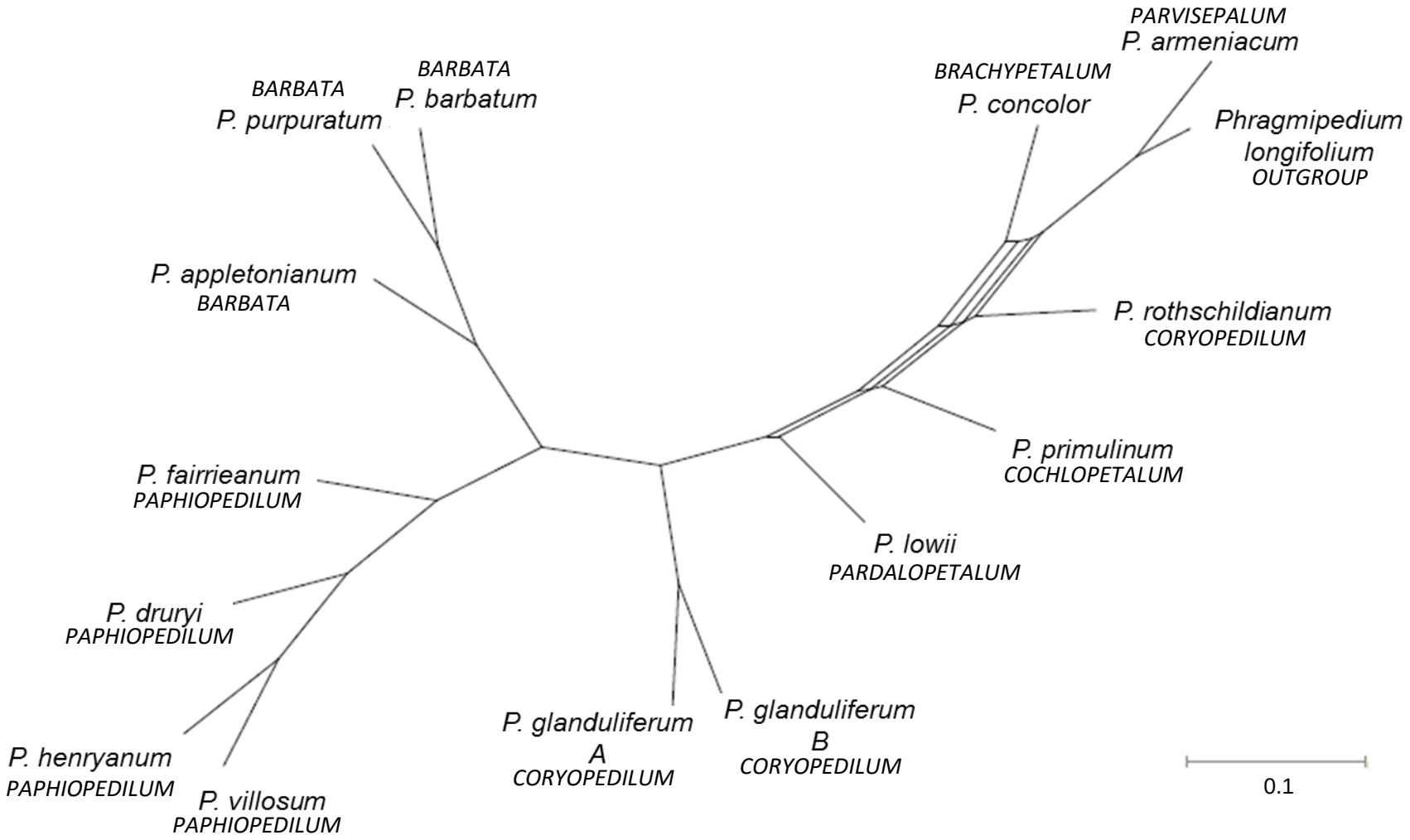

Figure 4.6 Filtered supernetwork showing relationships present in 10% of the bootstrap trees from the RE tree in Fig. 4.5 visualising conflicting splits that are potentially due to hybridisation or incomplete lineage sorting.

CHAPTER 5
GENERAL CONCLUSION

5.1 What is driving diversification in *Paphiopedilum* section *Barbata*?

Evolutionarily young groups such as *Paphiopedilum* section *Barbata* present an opportunity to study early stages of speciation in orchids. A review of the habitat and biology of *Paphiopedilum* (see Chapter 1) suggests that species in section *Barbata* show only very subtle niche partitioning. All members of section *Barbata* are wet forest terrestrials that grow in shade on decaying leaf litter, with overlapping flowering seasons and utilise the Syrphidae hoverflies as pollinators. Thus, I propose that speciation in section *Barbata* is largely driven by the availability of suitable wet forest habitats with radiations occurring when new habitat becomes available for colonisation.

The young taxa (i.e. <3.5 Myr, see Chapter 3) from islands that have only recently been colonised by section *Barbata* offer clues into the mechanisms that may be operating in the early stages of speciation in this group. For example, the closely allied *P. papuanum* and *P. violascens* from Papua New Guinea and *P. bougainvilleanum* and *P. wentworthianum* from the Solomon Islands (Fig. 5.1) provide some insights. From the cloned *Xdh* data presented in Chapter 3, it is suspected that these species may have only recently colonized these islands (section 3.5.1) and the data further suggest that what is being observed is something that can best be described as an early stage 'incomplete ring species' scenario i.e. taxa showing a morphological gradient occurring over a geographical range, caused by gradual west to east dispersal of taxa but where the sequence data do not yet enable clear relationships to be determined.

Similar scenarios likely took place among older taxa in continental Asia and on the Sunda shelf landmasses. However, the greater age of these taxa (as suggested from the cloned *Xdh* data, Chapter 3, section 3.5.1) means that these 'ring species' would likely have been 'broken' i.e. periodically interrupted by climatic changes brought about by ice-ages. Suitable wet forest habitat areas would have likely expanded, contracted and shifted in response to the changing climates. The effect of climate would have been especially pronounced in the Sunda shelf islands of Borneo, Sumatra and Java which become reconnected as a single contiguous landmass at the height of each glacial maximum.

Shifting distribution would also bring previously allopatric taxa into contact, especially at the high sea-levels when much of Sunda shelf diversity became concentrated in refugia. Over time, recurring hybridisation between sympatric taxa would result in the formation of syngameons i.e. groupings of genetically related yet morphologically distinct entities. Examples of section *Barbata* syngameons include *P. superbiens* and *P. tonsum* in Sumatra (Fig. 5.2A) and *P. callosum* and *P. appletonianum* in southern Indochina (Thailand) (Fig. 5.2B). Over longer periods, it is predicted that hybridisation may eventually lead to the original entities within the old syngameons losing morphological distinctiveness and becoming recognised as a single taxonomic entity. Evidence for the hybrid origin of the taxon only becomes apparent when phylogenetic investigation reveals conflicting signals from the ancestral units. Certainly, from the conflicting phylogenetic trees generated from the DNA sequence data given in chapters 2 and 3, I propose that *P. lawrenceanum* from Borneo, for example, may be an ancient syngameon of *P. barbatum* and *P. callosum* (Fig. 5.3). Currently, the ranges of *P. barbatum* and *P. callosum* overlap in the northern part of the Malaysian Peninsula (the states of Kedah, Perlis and Penang). However, given the geographical history of region, it is possible that sympatric populations may have existed elsewhere on the exposed Sunda shelf, the remnants of which are captured in the refugia on Borneo where it is recognised as the taxon *P. lawrenceanum*.

Although empirical evidence is currently lacking, it is suggested that chromosomal incompatibilities may be responsible for maintaining the species integrity of *P. hookerae* and *P. sangii*, which interestingly show no sign of inter-taxon hybridisation despite being sympatric with many other taxa in section *Barbata* (Chapter 2). As these taxa have a karyotype (2*n*=28) not found elsewhere in section *Barbata*, it is speculated that chromosomal changes, such as large or multiple inversions that serve as barriers to hybridisation, may have occurred in their common ancestor during the divergence of these taxa from the rest of section *Barbata*. In addition, large genome size differences between taxa may also serve as a barrier to hybridisation.

5.2 Future work

The findings from Chapter 2 and 3 have provided evidence that diversification in *Paphiopedilum* is driven by hybridisation and biogeographical vicariance and provided novel insights into the phylogenetic relationships both within taxa belonging to section *Barbata* and at the subgenus/section level of *Paphiopedilum*. However, the delimitation of certain key taxa and groups (e.g. *P. rungsuriyanum*, and the *P. appletonianum-bullenianum* alliance), and section *Cochlopetalum*, remain unresolved. Additional information from multiple independent

fast-evolving markers e.g. repeat clusters (building on the results obtained in Chapter 4, section 4.5.3), microsatellites, candidate adaptive genes and epigenetic alterations (Paun et al. 2011) will hopefully help to further clarify the phylogenetic relationships of the genus.

In Chapter 3, the biogeographical history of section *Barbata*, as inferred from the low-copy nuclear gene *Xdh*, suggests that the group originated on the Sunda shelf, underwent a west to east colonisation across Southeast Asia, experienced frequent genetic exchanges between the Sunda shelf landmasses, and has recently undergone radiations on the islands. However, this analysis is incomplete as it reflects only the evolutionary history of a single gene locus. Biogeographical reconstructions using multiple independent assorting loci are needed to give a more complete understanding of the biogeographical and evolutionary history of section *Barbata*.

The role of chromosomal evolution in orchid diversification is an interesting but largely overlooked field of study (Givnish et al. 2015) despite orchids displaying numerous polyploid-driven and polyploid-independent changes in chromosome number in many subfamilies and genera within Orchidaceae (Brandham 1999). The work presented in Chapter 4, suggests that *P. hookerae* and *P. sangii* may offer an interesting model for studying how chromosomal changes may preserve species integrity in the face of opportunities for hybridisation. Further work is clearly needed to identify the chromosomal changes occurring across the genus, and perhaps especially in *P. hookerae* and *P. sangii*. One approach would be to develop more chromosome-specific probes using the methods given in Chapter 4 which successfully identified the satellite sequence (*SatA*) as a promising chromosome-specific probe since it produced distinct chromosomal bands following fluorescence *in situ* hybridization (FISH) (see Fig. 4.3and 4.4). By extending this approach there is the potential to develop further chromosomes-specific probes so that particular chromosomes, and perhaps also chromosomal regions, can be tracked in different species by FISH. Potential sequences already identified in chapter (e.g. *SatB*, *SatG*) are suggested as the first sequences that could be tested.

The identification of suitable flow cytometry (FCM) buffers in this study has opened up the avenue for obtaining new FCM-based estimates in the tessellated-leaf *Paphiopedilum* which were previously regarded as recalcitrant. Thus, it is now possible to expand and update the existing body of genome size data in this group. These developments, in addition to an increasing number of *Paphiopedilum* HiSeq datasets produced from other on-going studies (Lee Yung-I and Rene Sumlders pers. comm.), present an exciting opportunity to further develop *Paphiopedilum* as an orchid model for studying genome evolution.

Figures

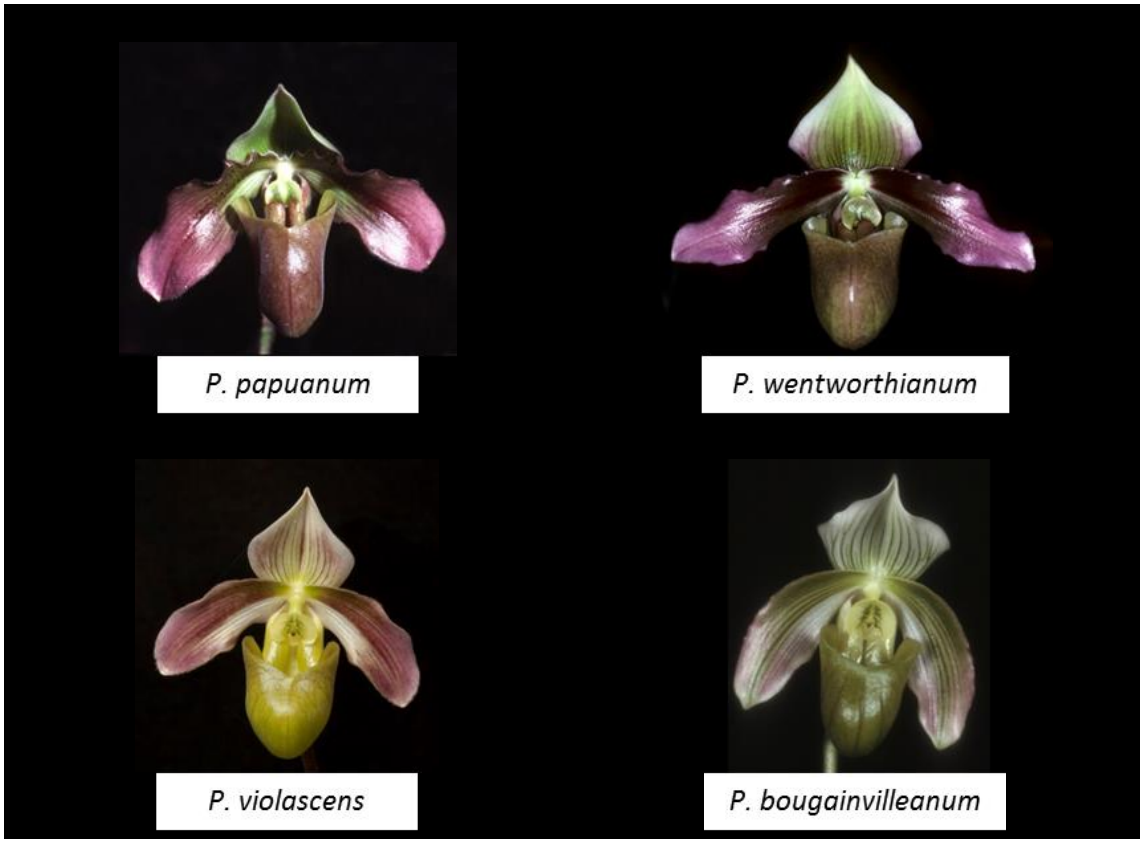

Figure 5.1 *Paphiopedilum papuanum*, *P. violascens*, *P. bougainvilleanum* and *P. wentworthianum*. The morphological gradient observed here suggests these taxa represent the early stages of a 'ring species'. Photo credits: *P. papuanum* (Jenny R), *P. wentworthianum* (Cribb PJ), *P. violascens* (Schuiteman A) & *P. bougainvilleanum* (Levy J).

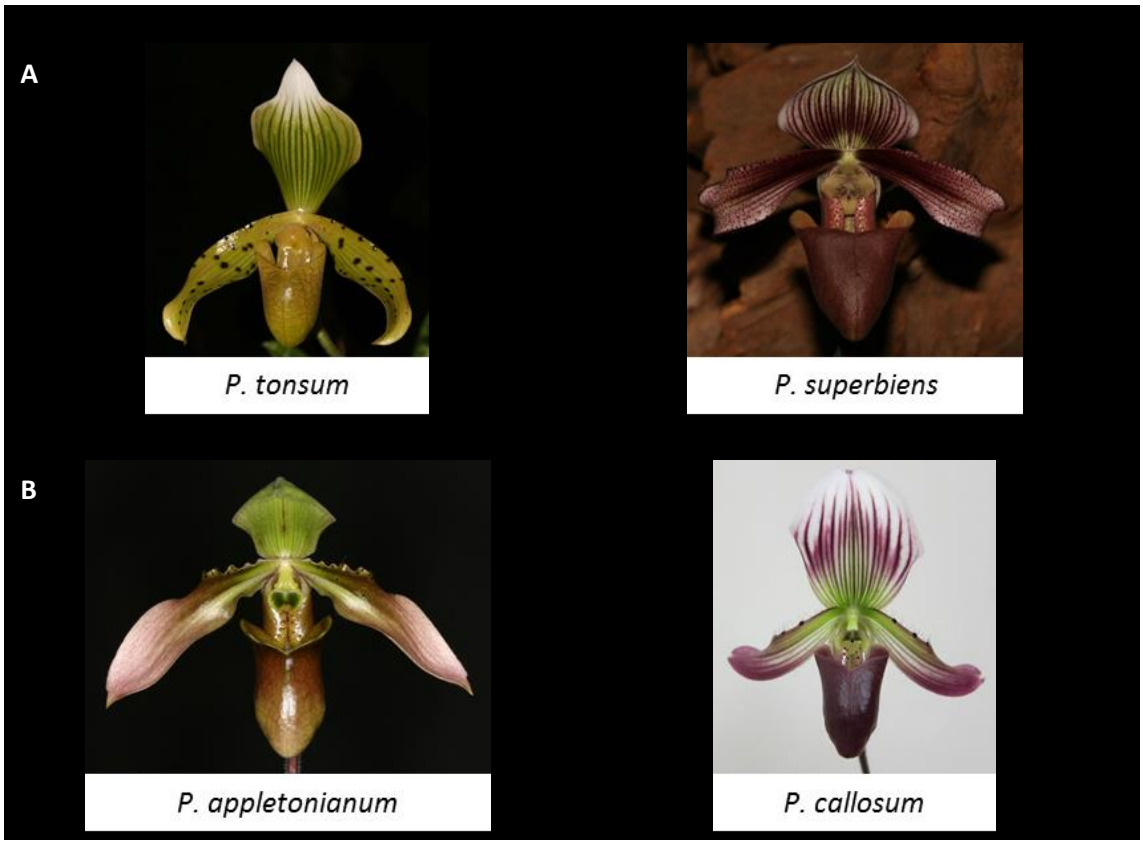

Figure 5.2 (A) *Paphiopedilum tonsum* and *P. superbiens* from Sumatra, and (B) *P. appletonianum and P. callosum* from South Indochina (Thailand). Examples of syngameons, genetically related but morphologically distinct groupings, in section *Barbata*. Photo credits: *P. tonsum* (Jenny R), *P. superbiens* (Jenny R), *P. appletonianum* (Schlumberger B) & *P. callosum* (Yap JW).

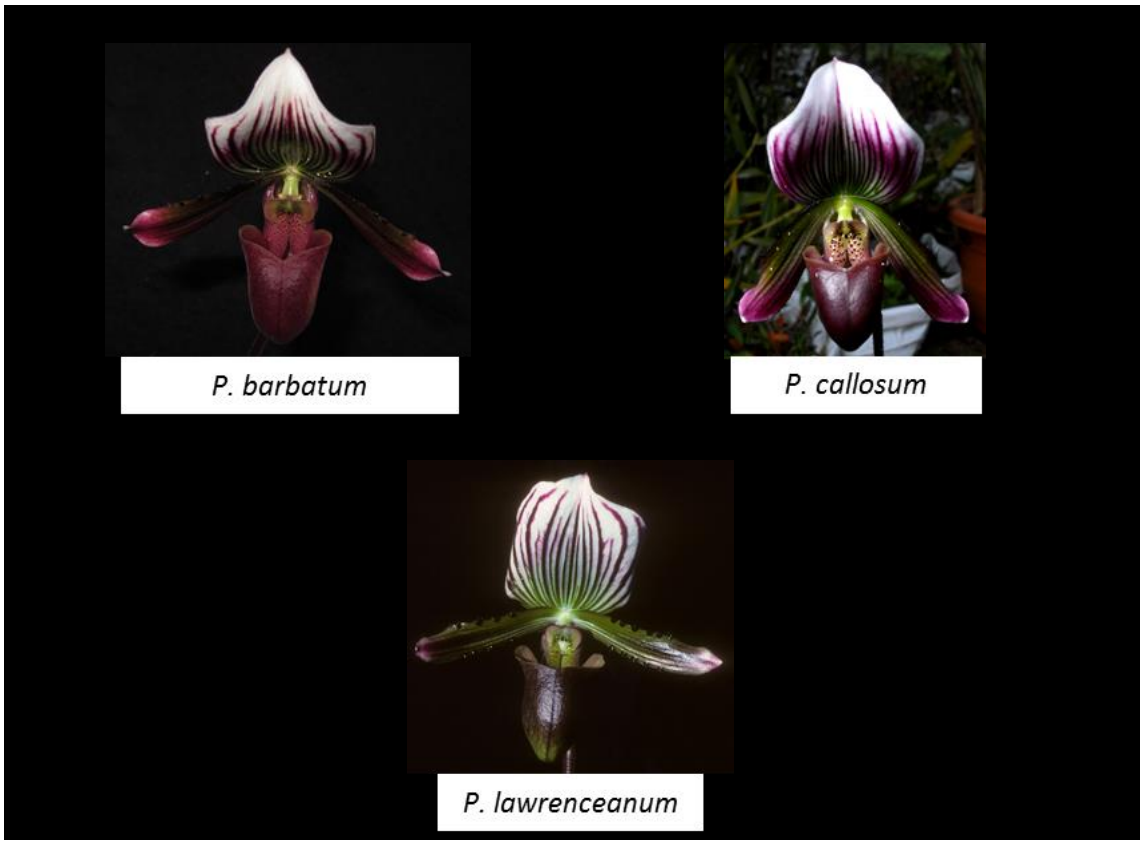

Figure 5.3 *Paphiopedilum barbatum*, *P. callosum* and *P. lawrenceanum*. Phylogenetic signals suggest that *P. lawrenceanum* arose from past hybridisation between *P. barbatum* and *P. callosum*. Photo credits: *P. barbatum* (Yap JW), *P. callosum* (Bryne PO) & *P. lawrenceanum* (Jenny R).

References

Abraham A, Ninan CA. **1954**. The chromosomes of *Ophioglossum reticulatum* L. *Current Science* **23**: 213–214.

Ågren JA, Wright SI. **2011**. Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? *Chromosome Research* **19**: 777–86.

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Dj L. **1997**. Gapped BLAST and PSI- BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Ambrozová K, Mandáková T, Bures P, Neumann P, Leitch IJ, Koblízková A, Macas J, Lysak MA. **2011**. Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. *Annals of Botany* **107**: 255–68.

Athipunyakom P, Manoch L, Piluek C. **2004**. Isolation and identification of mycorrhizal fungi from eleven terrestrial orchids. *The Kasetsart Journal* **38**: 216–228.

Atwood JT. **1984**. The relationships of the slipper orchids (subfamily Cypripedioideae, Orchidaceae). *Selbyana* **7**: 129–247.

Atwood JT. **1985**. Pollination of *Paphiopedilum rothschildianum*: brood-site deception. *National Geographic Research* **1**: 247–254.

Averyanov L V, Cribb P, Ke Loc P, Tien Hiep N. **2003**. *Slipper Orchids of Vietnam*. Kew, Richmond, Surrey: Royal Botanic Gardens.

Averyanov L V, Gruss O, Canh CX, Loc PK, Dang B, Hiep NT. **2010**. *Paphiopedilum canhii* - a new species from Northern Vietnam. *Orchids* **79**: 288–290.

Averyanov L V, Pham VT, Loc PK, Hiep NT, Canh CX, Vinh, Nguyen Quang Tien Hieu N. **2011**. "*Paphiopedilum canhii* – from Discovery to Extinction." *Orchid Planet* **N 24**: 16–44.

de Azkue D, Martínez A. **1988**. DNA content and chromosome evolution in the shrubby *Oxalis*. *Genome* **30**: 52–57.

Baack EJ, Whitney KD, Rieseberg LH. **2005**. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *The New Phytologist* **167**: 623–30.

Bänziger H. **1994**. Studies on the natural pollination of three species of wild lady-slipper orchids (*Paphiopedilum*) in Southeast Asia. In: Pridgeon A, ed. *Proceedings of the 14th World Orchid Conference*. Edinburgh: HMSO, 201–202.

Bänziger H. **1996**. The mesmerizing wart: the pollination strategy of epiphytic lady slipper orchid *Paphiopedilum villosum* (Lindl.) Stein (Orchidaceae). *Botanical Journal of the Linnean Society* **121**: 59–90.

Bänziger H. **2002**. Smart alecks and dumb flies: natural pollination of some wild lady slipper orchids (*Paphiopedilum* spp., Orchidaceae). In: Clark J,, In: Elliott W,, In: Tingley G,, In: Biro J, eds. *Proceedings of the 16th World Orchid Conference Vancouver*. Vancouver: Vancouver Orchid Society, 165–169.

Bänziger H, Pumikong S, Srimuang K. **2012**. The missing link: bee pollination in wild lady slipper orchids *Paphiopedilum thaianum* and *P. niveum* (Orchidaceae) in Thailand. *Mitteilungen der Schweizerischen …* **70**: 1–26.

Bänziger H, Sun H, Luo Y-B. **2008**. Pollination of wild lady slipper orchids *Cypripedium yunnanense* and *C. flavum* (Orchidaceae) in south-west China: why are there no hybrids? *Botanical Journal of the Linnean Society* **156**: 51–64.

Bennett MD, Bhandol P, Leitch IJ. **2000**. Nuclear DNA amounts in angiosperms and their

modern uses - 807 new estimates. *Annals of Botany* **86**: 859–909.

Bennett MD, Leitch IJ. **1997**. Nuclear DNA amounts in angiosperms - 583 new estimates. *Annals of Botany* **80**: 169–196.

Bennett MD, Leitch IJ. **2012**. Plant DNA C-values database (release 6.0, Dec. 2012). http://www.kew.org/cvalues/.

Bennett MD, Leitch IJ, Price HJ, Johnston JS. **2003**. Comparisons with *Caenorhabditis* (100 Mb) and *Drosophila* (175 Mb) using flow cytometry show genome size in *Arabidopsis* to be 157 Mb and thus 25 % larger than the *Arabidopsis* genome initiative estimate of 125 Mb. *Annals of Botany* **91**: 547–557.

Bennetzen JL, Wang H. **2014**. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annual Review of Plant Biology* **65**: 505–30.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. **2013**. GenBank. *Nucleic Acids Research* **41**: 36–42.

Braem G, Chiron G. **2003**. *Paphiopedilum*. Voreppe: Tropicalia.

Braem G, Gruss O. **2011**. *Paphiopedilum* subgenus *Megastaminodium* Braem & Gruss, a new subgenus to accommodate Paphiopedilum canhii. *Orchid Digest* **3**: 164.

Brandham P. **1999**. Cytogenetics. In: Pridgeon AM,, In: Cribb PJ,, In: Chase MW,, In: Rasmussen FN, eds. *Genera Orchidacearum Volume 1*. New York: NY: Oxford University Press, 67–80.

Brummitt RK. **2001**. *World Geographical Scheme for Recording Plant Distributions Edition 2*. Pittsburgh: Hunt Institute for Botanical Documentation, Carnegie-Mellon University.

de Bruyn M, Stelbrink B, Morley RJ, Hall R, Carvalho GR, Cannon CH, van den Bergh G, Meijaard E, Metcalfe I, Boitani L, Maiorano L, Shoup R, von Rintelen T. **2014**. Borneo and Indochina are Major Evolutionary hotspots for Southeast Asian biodiversity. *Systematic Biology* **63**: 879–901.

Butlin RK. **1993**. Species evolution - the role of chromosome change - King, M. *Nature* **4**: 27.

Cannon CH, Morley RJ, Bush ABG. **2009**. The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 11188–11193.

Chen SC, Cannon CH, Kua CS, Liu JJ, Galbraith DW. **2014**. Genome size variation in the Fagaceae and its implications for trees. *Tree Genetics and Genomes* **10**: 977–988.

Chen WH, Kao YL, Tang CY, Tsai CC, Lin TY. **2013**. Estimating nuclear DNA content within 50 species of the genus *Phalaenopsis* Blume (Orchidaceae). *Scientia Horticulturae* **161**: 70–75.

Chochai A, Leitch IJ, Ingrouille MJ, Fay MF. **2012**. Molecular phylogenetics of *Paphiopedilum* (Cypripedioideae; Orchidaceae) based on nuclear ribosomal ITS and plastid sequences. *Botanical Journal of the Linnean Society* **170**: 176–196.

Chung KS, Weber JA, Hipp AL. **2011**. Dynamics of chromosome number and genome size variation in a cytogenetically variable sedge (*Carex scoparia* var. *scoparia*, Cyperaceae). *American Journal of Botany* **98**: 122–129.

Conran JG, Bannister JM, Lee DE. **2009**. Earliest orchid macrofossils: Early Miocene Dendrobium and Earina (Orchidaceae: Epidendroideae) from New Zealand. *American Journal of Botany* **96**: 466–474.

Cox A V, Abdelnour GJ, Bennett MD, Leitch IJ. **1998**. Genome size and karyotype evolution in the slipper orchids (Cypripedioideae: Orchidaceae). *American Journal of Botany* **85**: 681–687.

Cox A V, Eon AMPR, Albert VA, Chase MW. **1997**. Phylogenetics of the slipper orchids (Cypripedioideae, Orchidaceae): nuclear rDNA ITS sequences. *Plant Systematics and Evolution*

208: 197–223.

Coyne JA, Orr HA. **2004**. *Speciation*. Sunderland, MA: Sinauer Associates.

Crayn DM, Costion C, Harrington MG. **2015**. The Sahul-Sunda floristic exchange: dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics. *Journal of Biogeography* **42**: 11–24.

Cribb P. **1998**. *The Genus Paphiopedilum*. Natural History Publications (Borneo), Malaysia.

Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novak P, Piednoel M, Weiss-Schneeweiss H, Leitch AR. **2015**. Genomic repeat abundances contain phylogenetic signal. *Systematic Biology* **64**: 112–126.

Dodsworth S, Chase MW, Särkinen T, Knapp S, Leitch AR. **2015**. Using genomic repeats for phylogenomics: a case study in wild tomatoes (*Solanum* section *Lycopersicon*: Solanaceae). *Biological Journal of the Linnean Society*: doi 10.1111/bij.12612.

Doležel J, Bartoš J, Voglmayr H, Greilhuber J, Thomas RA. **2003**. Nuclear DNA content and genome size of trout and human (multiple letters). *Cytometry Part A* **51**: 127–129.

Doležel J, Greilhuber J, Lucretti S, Meister AE, Lysak MA, Nardi L, Obermayer R. **1998**. Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Annals of Botany* **82**: 17–26.

Doyle JJ, Doyle JL. **1987**. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin of the Botanical Society of America* **19**: 11–15.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. **2006**. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: 699–710.

Drummond AJ, Suchard MA, Xie D, Rambaut A. **2012**. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–73.

Ebert D, Peakall R. **2009**. A new set of universal de novo sequencing primers for extensive coverage of noncoding chloroplast DNA: New opportunities for phylogenetic studies and cpSSR discovery. *Molecular Ecology Resources* **9**: 777–783.

Ebihara A, Ishikawa H, Matsumoto S, Lin SJ, Iwatsuki K, Takamiya M, Watano Y, Ito M. **2005**. Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the *Vandenboschia radicans* complex (Hymenophyllaceae) in Japan and adjacent areas. *American Journal of Botany* **92**: 1535–1547.

Edens-Meier R, Luo Y-B, Pemberton RW, Bernhardt P. **2015**. Pollination and floral evolution of slipper orchids (subfamily Cypripedioideae). In: Edens-Meier R,, In: Bernhardt P, eds. *Darwin's Orchids: Then and Now*. Chicago and London: The University of Chicago Press, 265–287.

Felsenstein J. **1985**. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Felsenstein J. **1987**. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**: 124–138.

Fitch WM. **1971**. Toward defining the course of evolution: Minimum change for a specific tree topology. *Systematic Biology* **20**: 406–416.

Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Temsch EM, Greilhuber J, Müller KF, Heubl G. **2014**. Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Annals of Botany* **114**: 1651–63.

Futuyama DJ, Mayer GC. **1980**. Non-allopatric speciation in animals. *Systematic Zoology* **29**: 254–271.

Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Iles WJD, Clements MA, Arroyo

MTK, Leebens-Mack J, Endara L, Kriebel R, Neubig KM, Whitten WM, Williams NH, Cameron KM. **2015**. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society B* **282**: DOI: 10.1098/rspb.2015.1553.

Goloboff PA, Farris JS, Nixon KC. **2008**. TNT, a free program for phylogenetic analysis. *Cladistics* **24**: 774–786.

Górniak M, Paun O, Chase MW. **2010**. Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, Xdh: Congruence with organellar and nuclear ribosomal DNA results. *Molecular Phylogenetics and Evolution* **56**: 784–795.

Górniak M, Szlachetko DL, Kowalkowska AK, Bohdanowicz J, Canh CX. **2014**. Taxonomic placement of *Paphiopedilum canhii* (Cypripedioideae; Orchidaceae) based on cytological, molecular and micromorphological evidence. *Molecular Phylogenetics and Evolution* **70**: 429–441.

Grant V. **1981**. *Plant Speciation*. Columbia: Columbia University Press.

Graur D, Martin W. **2004**. Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision. *Trends in Genetics* **20**: 80–86.

Greilhuber J, Temsch EM, Loureiro J. **2007**. Nuclear DNA Content Measurement. In: Dolezel J,, In: Greilhuber J,, In: Suda J, eds. *Flow Cytometry with Plant Cells*. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA, 67–101.

Guo Y-Y, Luo Y-B, Liu Z-J, Wang X-Q. **2012**. Evolution and biogeography of the slipper orchids: Eocene vicariance of the conduplicate genera in the old and new world tropics. *PLoS One* **7**: e38788.

Guo Y-Y, Luo Y-B, Liu Z-J, Wang X-Q. **2015**. Reticulate evolution and sea-level fluctuations together drove species diversification of slipper orchids (*Paphiopedilum*) in Southeast Asia. *Molecular Ecology* **24**: 2838–2855.

Hall R. **2002**. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences* **20**: 353–431.

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF. **2006**. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. *Genome Research* **16**: 1252–61.

Iles WJD, Smith SY, Gandolfo MA, Graham SW. **2015**. Monocot fossils suitable for molecular dating analyses. *Botanical Journal of the Linnean Society* **178**: 346–374.

Inda LA, Pimentel M, Chase MW. **2010**. Chalcone synthase variation and phylogenetic relationships in *Dactylorhiza* (Orchidaceae). *Botanical Journal of the Linnean Society* **163**: 155–165.

Kamemoto H, Sagarik R, Dieutrakul S. **1963**. Karyotypes of *Paphiopedilum* species of Thailand. *The Kasetsart Journal* **3**: 69–78.

Karasawa K. **1979**. Karyomorphological studies in *Paphiopedilum*, Orchidaceae. *Bulletin of the Hiroshima Botanical Garden* **2**: 1–149.

Karasawa K. **1980**. Karyomorphological studies in *Phragmipedium*, Orchidaceae. *Bulletin of the Hiroshima Botanical Garden* **3**: 1–49.

Karasawa K. **1982**. Karyomorphological studies on four species of *Paphiopedilum*. *Bulletin of the Hiroshima Botanical Garden* **5**: 70–79.

Karasawa K. **1986**. Karyomorphological studies on nine taxa of *Paphiopedilum*, Orchidaceae. *Bulletin of the Hiroshima Botanical Garden* **8**: 23–42.

Karasawa K, Aoyama M. **1980**. Karyomorphological studies on two species of *Paphiopedilum*.

Bulletin of the Hiroshima Botanical Garden **3**: 69–74.

Karasawa K, Aoyama M. **1988**. Karyomorphological studies on two species of *Paphiopedilum*. *Bulletin of the Hiroshima Botanical Garden* **10**: 1–6.

Karasawa K, Aoyama M, Kamimura T. **1997**. Karyomorphological studies on five rare species of Paphiopedilum, Orchidaceae. *Annals of the Tsukuba Botanical Garden* **16**: 29–39.

Karasawa K, Saito K. **1982**. A revision of the genus *Paphiopedilum* (Orchidaceae). *Bulletin of the Hiroshima Botanical Garden* **5**: 1–69.

Karasawa K, Tanaka R. **1980**. C-banding study on centric fission in the chromosome of Paphiopedilum. *Cytologia* **45**: 97–102.

Karasawa K, Tanaka R. **1981**. A revision of chromosome number in some hybrids of Paphiopedilum. *Bulletin of the Hiroshima Botanical Garden* **4**: 1–8.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. **2012**. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–9.

Kelly LJ, Leitch IJ. **2011**. Exploring giant plant genomes with next-generation sequencing technology. *Chromosome Research* **19**: 939–53.

Kelly LJ, Renny-byfield S, Pellicer J, Nov P, Neumann P, Lysak MA, Day PD, Berger M, Fay MF, Nichols RA, Leitch AR, Leitch IJ, Kelly LJ. **2015**. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. *New Phytologist*: doi: 10.1111/nph.13471.

King M. **1993**. *Species Evolution: the role of chromosome chamge*. Cambridge: Cambridge University Press.

Kirkpatrick M, Barton N. **2006**. Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–34.

Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A. **2010**. Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. *The New Phytologist* **186**: 148–60.

Lan T, Albert VA. **2011**. Dynamic distribution patterns of ribosomal DNA and chromosomal evolution in *Paphiopedilum*, a lady's slipper orchid. *BMC Plant Biology* **11**: 126.

Lee Y-I, Chang F-C, Chung M-C. **2011**. Chromosome pairing affinities in interspecific hybrids reflect phylogenetic distances among lady's slipper orchids (*Paphiopedilum*). *Annals of Botany* **108**: 113–121.

Leitch IJ, Beaulieu JM, Chase MW, Leitch AR, Fay MF. **2010**. Genome size dynamics and evolution in Monocots. *Journal of Botany* **2010**: 1–18.

Leitch IJ, Kahandawala I, Suda J, Hanson L, Ingrouille MJ, Chase MW, Fay MF. **2009**. Genome size diversity in orchids: consequences and evolution. *Annals of Botany* **104**: 469–81.

Liu Z, Chen L, Zhou Q. **2010**. *The reproduction strategy of Paphiopedilum armeniacum (Orchidaceae)*. Nova Science Publishers, Inc.

Liu Z, Liu K, Chen L, Lei S, Li L, Shi X, Huang L. **2006**. Conservation ecology of endangered species *Paphiopedilum armeniacum* (Orchidaceae). *Acta Ecologica Sinica* **26**: 2791–2799.

Liu ZJ, Zhang JY, Ru ZZ, Lei SP, Chen LJ. **2004**. Conservation biology of *Paphiopedilum purpuratum* (Orchidaceae). *Biodiversity Science* **12**: 509–516.

Lohman DJ, de Bruyn M, Page T, von Rintelen K, Hall R, Ng PKL, Shih H-T, Carvalho GR, von Rintelen T. **2011**. Biogeography of the Indo-Australian Archipelago. *Annual Review of Ecology, Evolution, and Systematics* **42**: 205–226.

Marques I, Nieto Feliner G, Martins-Loução MA, Fuertes Aguilar J. **2012**. Genome size and base composition variation in natural and experimental *Narcissus* (Amaryllidaceae) hybrids. *Annals of Botany* **109**: 257–64.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. **2015**. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* **1**: 1–5.

Matzke MA, Mosher RA. **2014**. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics* **15**: 394–408.

Maumus F, Quesneville H. **2014**. Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. *PloS One* **9**: e94101.

McQuaid HA. **1949**. The cytology of *Paphiopedilum* Maudiae Hort. *Annals of the Missouri Botanical Garden* **36**: 433–473.

Miller KG, Mountain GS, Wright JD, Browning JV. **2011**. Sea level and ice volume variations. *Oceanography* **24**: 40–53.

Miller MA, Pfeiffer W, Schwartz T. **2010**. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop, GCE 2010*.

Morley RJ. **2000**. *Origin and Evolution of Tropical Rain Forests*. Chichester, UK: Wiley.

Morley RJ. **2012**. A review of the Cenozoic palaeoclimate history of Southeast Asia. In: Gower D,, In: Johnson KG,, In: B.R. R,, In: J. R,, In: L. R,, In: Williams ST, eds. *Biotic Evolution and Environmental Change in Southeast Asia*. Cambridge: Cambridge University Press, 79–114.

Nakazato T, Barker MS, Rieseberg LH, Gastony GJ. **2008**. *Biology and Evolution of Ferns and Lycophytes* (TA Ranker and CH Haufler, Eds.). Cambridge: Cambridge University Press.

Navarro A, Barton NH. **2003**. Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. *Evolution* **57**: 447–459.

Neubig KM, Whitten WM, Carlsward BS, Blanco M a., Endara L, Williams NH, Moore M. **2009**. Phylogenetic utility of ycf1 in orchids: a plastid gene more variable than matK. *Plant Systematics and Evolution* **277**: 75–84.

Noor MA, Grams KL, Bertucci LA, Reiland J. **2001**. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 12084–12088.

Novák P, Hřibová E, Neumann P, Koblížková A, Doležel J, Macas J. **2014**. Genome-wide analysis of repeat diversity across the family musaceae. *PLoS One* **9**: e98918.

Novák P, Neumann P, Macas J. **2010**. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **11**: 378.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. **2013**. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29**: 792–793.

Nylander JA. **2004**. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Obermayer R, Leitch IJ, Hanson L, Bennett MD. **2002**. Nuclear DNA C-values in 30 species double the familial representation in Pteridophytes. *Annals of Botany* **90**: 209–217.

Paun O, Bateman RM, Fay MF, Luna JA, Moat J, Hedrén M, Chase MW. **2011**. Altered gene expression and ecological divergence in sibling allopolyploids of *Dactylorhiza* (Orchidaceae). *BMC Evolutionary Biology* **11**: 113.

Pellicer J, Fay MiF, Leitch IJ. **2010**. The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society* **164**: 10–15.

Pellicer J, Kelly LJ, Leitch IJ, Zomlefer WB, Fay MF. **2014**. A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. *The New Phytologist* **201**: 1484–97.

Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O. **2006**. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. *Genome Research* **16**: 1262–9.

Raes N, Cannon CH, Hijmans RJ, Piessens T, Saw LG, van Welzen PC, Slik JWF. **2014**. *Historical distribution of Sundaland's Dipterocarp rainforests at Quaternary glacial maxima*.

Rambaut A, Suchard MA, Xie W, Drummond AJ. **2013**. MCMC Trace Analysis Tool. : Tracer v1.6, Available from http://beast.bio.ed.ac.

Ramírez SR, Gravendeel B, Singer RB, Marshall CR, Pierce NE. **2007**. Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature* **448**: 1042–1045.

Rasmussen HN. **1995**. *Terrestrial orchids: from seed to mycotrophic plant*. Cambridge: Cambridge University Press.

Ree RH, Moore BR, Webb CO, Donoghue MJ. **2005**. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution; international journal of organic evolution* **59**: 2299–2311.

Ree RH, Smith SA. **2008**. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* **57**: 4–14.

Richardson JE, Costion CM, Muellner AN. **2012**. The Malesian floristic interchange: plant migration patterns across Wallace's Line. In: Gower D,, In: Johnson K,, In: Richardson J,, In: Rosen B,, In: Rüber L,, In: Williams S, eds. *Biotic Evolution and Environmental Change in Southeast Asia*. Cambridge: Cambridge University Press, 138–163.

Rieseberg LH. **2001**. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution* **16**: 351–358.

Robinson TJ. **1995**. M. King., Species evolution: the role of chromosome change. *Systematic Biology* **44**: 578–580.

Shi J, Cheng J, Luo D, Shangguan F-Z, Luo Y-B. **2007**. Pollination syndromes predict brood-site deceptive pollination by female hoverflies in *Paphiopedilum dianthum* (Orchidaceae). *Acta Phytotaxonomica Sinica* **45**: 551–560.

Shi J, Luo Y-B, Bernhardt P, Ran J-C, Liu Z-J, Zhou Q. **2009**. Pollination by deceit in *Paphiopedilum barbigerum* (Orchidaceae): a staminode exploits the innate colour preferences of hoverflies (Syrphidae). *Plant Biology* **11**: 17–28.

Shi J, Luo Y-B, Cheng J, Shangguan F-Z, Deng Z-H. **2009**. The pollination of *Paphiopedilum hirsutissimum*. *Orchid Review* **117**: 78–81.

Sonnhammer ELL, Durbin R. **1995**. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**: GC1–GC10.

Spakman W, Hall R. **2010**. Surface deformation and slab–mantle interaction during Banda arc subduction rollback. *Nature Geoscience* **3**: 562–566.

Stebbins GL. **1971**. *Chromosomal evolution in higher plants*. London: Arnold.

Stelbrink B, Albrecht C, Hall R, Rintelen T Von. **2012**. The biogeography of Sulawesi revisited: is there evidence for a vicariant origin of taxa on Wallace's " anomalous island "? *Evolution* **66**: 2252–2271.

Strasburg JL, Scotti-Saintagne C, Scotti I, Lai Z, Rieseberg LH. **2009**. Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Molecular*

Biology and Evolution **26**: 1341–1355.

Sun H, Mclewin W, Fay MF. **2001**. Molecular phylogeny of *Helleborus* (Ranunculaceae), with an emphasis on the East Asian-Mediterranean disjunction. *Taxon* **50**: 1001–1018.

Swofford DL. **2002**. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Trávníček P, Ponert J, Urfus T, Jersáková J, Vrána J, Hřibová E, Doležel J, Suda J. **2015**. Challenges of flow-cytometric estimation of nuclear genome size in orchids, a plant group with both whole-genome and progressively partial endoreplication. *Cytometry Part A*.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. **2012**. Primer3-new capabilities and interfaces. *Nucleic Acids Research* **40**: 1–12.

Vitte C, Panaud O. **2005**. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and Genome Research* **110**: 91–107.

Walsh JB. **1982**. Rate of accumulation of reproductive isolation by chromosomal rearrangements. *The American Naturalist* **120**: 510–532.

White MJD. **1978**. *Modes of Speciation*. San Francisco: W.H. Freeman and Company.

Yang Z, Rannala B. **1997**. Bayesian phylogenetic Monte Carlo method inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular Biology and Evolution* **14**: 717–724.

Yuan L, Yang ZL, Li S-Y, Hu H, Huang J-L. **2010**. Mycorrhizal specificity, preference, and plasticity of six slipper orchids from South Western China. *Mycorrhiza* **20**: 559–568.

Appendix

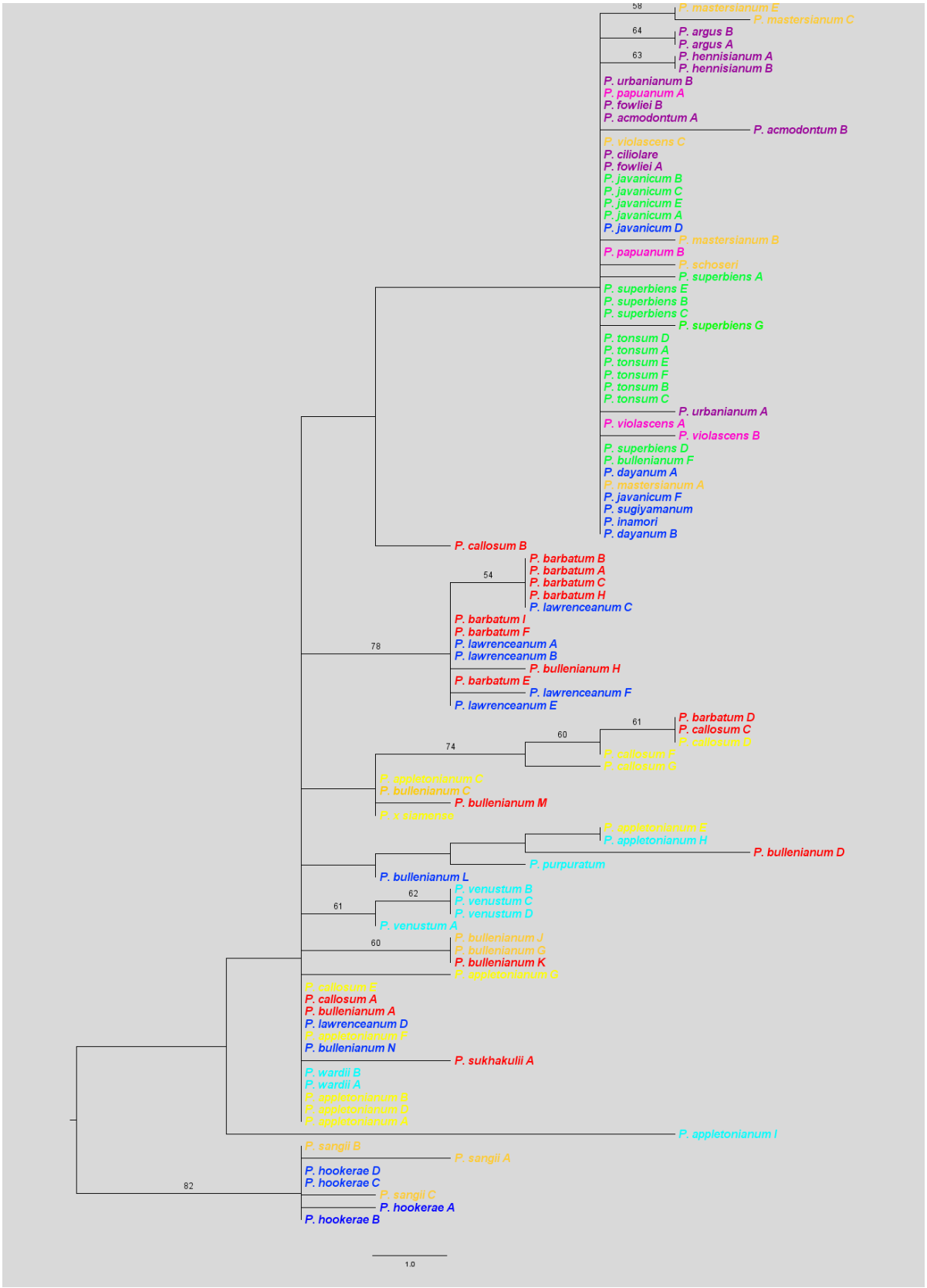

Figure A A tree showing the taxon relationships within section *Barbata* produced by MP analysis of the nuclear *Xdh* region. Numbers above branches indicate bootstrap (BP) percentages above 50. The colours of the tips indicate their origin and correspond to the geographical ranges displayed in Fig. 2.4.

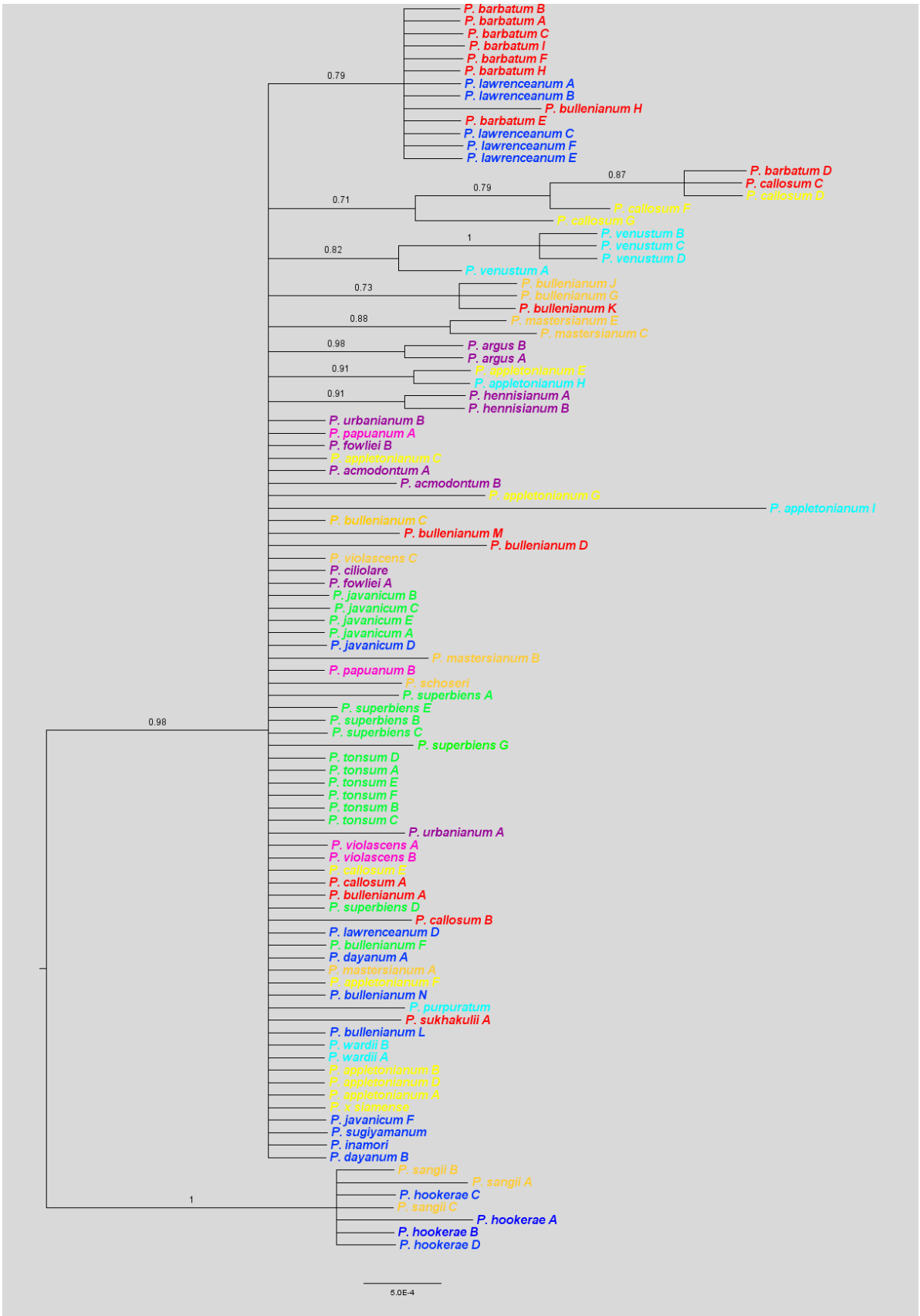

Figure B A tree showing the taxon relationships within section *Barbata* produced by Bayesian analysis of the nuclear *Xdh* region. Numbers above branches indicate posterior probability (PP) scores above 0.5. The colours of the tips indicate its origin and correspond to the geographical ranges displayed in Fig. 2.4.